

**Whitepaper.**

**Making the move  
to open source:  
Top tips for a successful  
SAS to R migration.**



# Contents.

INTRODUCTION	3
MODERN ANALYTIC PRACTICES	4
THE DRIVING FORCE FOR CHANGE	5
PROJECT EXECUTION	8
MIGRATION TO OPEN SOURCE: A PREMIER INC. CASE STUDY	10
CONCLUSION	12



# Introduction.

Commercial decision-making is increasingly the driver for change in modern analytic practice. One major decision is around open source vs. proprietary: whether you build your technical foundations on freely available software like R or Python and become part of the global community of contributors, or opt for an enterprise licensed product like SAS, which is developed and maintained centrally.

SAS has long been established as the leader in the commercial analytics space. However in recent years, this position has been increasingly challenged by the evolution - and growing adoption - of more flexible and cost-effective open source alternatives.

Making the decision to move away from SAS is often a consequence of a broader company-wide transformation around the use of data to drive strategic decision-making. And once the strategic decision is made, organisations need to face into the human and technical challenges of migration: from code conversion to culture.

This paper explores trends in migration, its associated challenges and experience-based advice for organisations about to embark on the switch to open source.

## About SAS.

SAS is a long standing, comprehensive platform widely used by companies across a range of industries. Easy to use, it offers the user a range of statistical functionalities and a straightforward GUI. SAS is easy to learn and backed by technical support. However, unlike open source technologies, the adoption of SAS commits organisations to license-based contracts and a centrally governed product roadmap.



# Modern analytic practices.

The visibility and remit of analytics has changed over the years as organisations increasingly understand that future success depends largely on the use of data to make optimal business decisions and drive efficiencies.

This may require a change from reactive statistical analyses around a narrow range of challenges to more proactive open source data science methods to drive business decisions in a repeatable and scalable manner. Implementing a holistic modern analytic approach often impacts working practices and workflows across the organisation, including the underlying technology platform.

	<b>Reactive.</b>	<b>Proactive.</b>
<b>Output</b>	Insight communicated via graphs and tables, with code as a by-product.	Repeatable code and apps with insight conveyed in richer interactive formats.
<b>Skills</b>	Individual skills around scripting and knowledge of specific analytic method.	Emphasis on team skills. A broader range of talent and empathy to engage business.
<b>Processes</b>	Internal analytic workflows focused on output of repetitive insight.	Engaged workflow clearly understood by the business to solve a range of challenges.
<b>Technology</b>	Platform and tools adept at repeating similar tasks based on statistical methods.	More flexible tools capable of a wider range of analytic approaches and richer outputs.



# The driving force for change.

There are a number of contributing factors to the decision to move from SAS to open source.



More portability for scaling solutions.



More flexibility for experimentation.



Access to richer capabilities.



Moving from reactive to proactive stats analysis.



Environment limitations.



Better alignment with business processes.



High/increasing licence costs.



Availability of technical skills.

## The desire to reduce annual licensing costs.

SAS is routinely used by large organisations thanks to its ease of adoption and formal support options, and is procured and delivered via commercial licensing contracts.

However as IT budgets have faced unprecedented pressure over the last few years, software licensing has been an obvious candidate for cost reduction. The immediacy of the saving makes open source options highly attractive - and a universal feature of the migration business case.



## The desire to adopt modern data science techniques & methodologies.

Another prevalent reason for migrating away from SAS to R or Python is the quest for the latest functionality and access to richer capabilities. Historically, it was considered that a specific statistical routine was not widely accepted until it was implemented in SAS. This sentiment is no longer evident - modern analysts and data scientists need access to cutting-edge models and approaches that are just not incorporated in SAS quickly enough.

SAS is a large piece of software designed to perform standard ETL and analytic processes in a repetitive manner. SAS is also used interactively, but challenges can occur when you experiment and innovate. Generally, SAS is perfectly suited to standard activities but can be painful when trying to do something the software was never designed to do - and businesses looking to get one step ahead of their competition by utilising the latest and greatest algorithms simply don't have the time to wait for the next SAS release.

Today there are a vast range of support options and products available that allow analytics teams to deploy R in production environments, giving organisations access to massive flexibility and scale. This evolution responds to an early market perception that only commercially licensed enterprise products could offer sufficient security, reliability and accountability, which has been repeatedly disproved in critical use cases in highly regulated sectors.





## The ability to formulate efficient data science practices with delivery-focused methodology.

A SAS to open source migration may first appear as a simple technical project. However, it can impact a range of technical and more importantly, cultural factors.

Imagine you take a Software Developer with a formal programming education and ask them to transition from Java to C#. They may initially be frustrated, but their core 'programming' capability will support their transition as they apply their skills to Java - so they can go from this base to C# and find familiar structures and concepts. A SAS user typically doesn't have a programming background, so they don't recognise they have an abstract ability to create code - all they know is SAS as the principal mechanism for doing their job. A migration from SAS to another system is a very worrying prospect for many long-term SAS users. Strong cultural objections to a migration are often driven from this perspective, so critical to success is an understanding of the context of the migration, and an empathy with the user community.

Migrations must therefore include a large amount of appropriate communication, training support and full enablement for any technology adoption to be successful.

Another cultural challenge is the fact that legacy SAS usage may be based around set coding standards and procedures which can create barriers when adopting new technologies. Similar standards and procedures are required for R or Python, so care must be taken to understand how R/Python code is to be developed, tested and deployed to end users.

## To entice the modern data scientist to join the organisation.

In recent years, the need to attract, resource and upskill data science teams has become far more pronounced. With the skills gap growing ever wider, employers need to take proactive measures to retaining leading talent.

R and Python are recognised leading technologies for data science today and have been widely adopted within many major organisations.

Universities have adopted R as their primary analytical programming language for mathematics, science and statistics courses meaning graduates are joining the workplace fully skilled and suitably ready to deploy their newfound skills. For GCSE age school children, Python coding is now on the National Curriculum, as part of the Computer Science course.

By contrast, SAS skills are declining as universities shift away from its teaching and unlike other languages, SAS has limited free online resources which means costly training. It is simply not cost-effective or sensible for any organisation to take people already skilled in R or Python and train them to use SAS.

Such grassroots formation in open source languages is fuelling the trend away from SAS. This is starting to take shape with the emergence of centralised data science and analytics capabilities in open source technologies across a much wider range of sectors.



# Project execution.

A big part of getting a PoC moving internally is building credibility.

As with any major projects, there are potential issues and pitfalls involved in converting from a single tool such as SAS. Typically, organisations have built up large amounts of legacy code over the years; much of this may be undocumented and untested meaning a conversion project can be a huge task. SAS teams have typically invested years of effort building large libraries of SAS macros and code. The move to open source is not something, therefore, which happens overnight.

In proposing a strategic move away from SAS, companies must carefully consider a variety of factors including re-training staff and crucially, the migration of what is typically a very large code base developed over many years by a variety of analysts. So how do you migrate 20 years' worth of SAS code to R/Python? And how do you begin tackling these challenges and delivering your migration to open source?







## Taking on the migration challenge.

Here are our 6 key components of a successful SAS migration project.



### Step 1. Strategy

Create an effective strategy with clearly defined business objectives (e.g. x% to convert, level of competency) and an evaluation of risk, alongside a well-informed change/communication plan with a clearly defined roadmap and success criteria.



### Step 2. Skills

Assess the team's competency in the chosen open source language. Build a strategic approach to learning and work with the team to align this with their personal development objectives. Augment off-the-shelf learning with custom courses to elevate and help retain talent.



### Step 3. Data

Migrate SAS data files held on local servers or stored in legacy databases into a more appropriate environment. With a digital transformation there are considerable benefits of migrating the data architecture to a more modern cloud based architecture alongside the overall code migration.



### Step 4. Platform

Moving from a SAS system to a R-centric or Python data platform gives you the ability to scale and flex in line with your analytic journey. Ensure it is fit for purpose in the modern analytic world with accessibility to rich functionality from a single environment in which practitioners can quickly import, clean, visualise and model data.



### Step 5. Process

Change workflows to represent new ways of working with the business, with the analytic team and with R/Python itself. Structure and measure success in a consistent way so that employees at all levels can see the data-driven programme at work.



### Step 6. Action

Convert legacy SAS code to R/Python and provide evidence and assurance to stakeholders around the newly migrated code base. Early adopters can help support testing, demonstrate measurable value and help identify any blockers.



# Migration to open source: a Premier Inc. case study.

## About Premier Inc.

Healthcare improvement company Premier Inc unites more than 4,400 US hospitals and approximately 225,000 other providers and organisations to transform healthcare. Named one of the World's Most Ethical Companies® 13 years in a row, Premier empowers its members to improve care quality and safely reduce cost with data-driven insight - and has created one of the most comprehensive databases of actionable data, clinical best practice and efficiency improvement strategies available worldwide.

## The challenge.

Internally, their 30-strong data consulting team were reliant on SAS for both research questions and customised data services. However, increasing SAS licence renewal rates and restrictive terms and conditions prompted the company to search for a more modern and cost-effective platform. After a period of consultation, Premier decided to adopt a faster, more capability rich open source R environment for their internal teams to accelerate innovation and attract new data science talent.





## The solution.

Working closely with Premier, we managed the strategic adoption, implementation and onboarding programme to land the Premier team successfully in the new RStudio Enterprise environment.

We delivered the agile programme in line with Premier's pace and direction, splitting the project into three workstreams: skills, code migration and enablement. Assessing the capabilities of the Premier team, we provided custom training in all aspects of R packages to drive faster insight, including ML, Shiny and web scraping. Our Data Scientists were engaged to migrate the legacy code for their three main SAS-coded frameworks to R and worked on a 1:1 basis to enable the Premier team to rapidly increase their coding proficiency in R.

## The outcome.

In making the transition from SAS to R, Premier has substantially reduced annual licensing renewal costs, gained a scalable solution for experimentation and innovation and adopted a data environment that is far more attractive to its current and future data science talent.

Premier's Data Scientists and Researchers are now able to create and manage existing frameworks, models, packages and consulting work products in R. As a result, the teams are more engaged and empowered, sharing knowledge and skills in a community of practice centered around R, which in turn makes it easier for them to find, recruit and retain talent with open source skills. The project delivered:

- All planned SAS licensing cost savings
- 40+ Data Scientists and researchers successfully migrated from existing SAS-based frameworks and trained in R in approximately 6 months
- A well-structured RStudio implementation
- The empathetic support and coaching required for onward success
- 100% talent retention.



# Conclusion.

There are many reasons for a migration away from SAS to open source, whether this is cost reduction, access to richer capabilities, more flexibility for experimentation or simply the availability of technical skills.

Frequently, however, a migration is a consequence of a broader data-driven transformation and a business-wide adoption of modern analytical practices. Whatever the drivers for the migration may be, the considerations and steps outlined above provide an essential framework for successful transition.

Whilst not all projects are the same, they need to be aligned to the needs of the team and business objectives. Ongoing evaluation and project management will ensure consistency in delivery and prioritisation, as well ensuring successful adoption.

## Next steps.

The creation of a practical strategy aligned to internal business objectives helps to shape your migration roadmap. Some essential elements might include:

- Your business objectives and success criteria
- A current state audit
- Your target state design
- A high level roadmap
- A POC to validate your benefits.

If however you are looking for an experienced SAS migration partner with the technical capability, performance and resource scale to accelerate your migration and help avoid or reduce SAS extension costs, get in touch:

[letstalk@ascent.io](mailto:letstalk@ascent.io).

