

Don't just survive. Thrive! A Learning-Based Strategy for a Modern Organization Centered Around SAS®

Jovan Marjanovic, ProWerk Consulting

ABSTRACT

SAS® is still the undisputed market leader in the commercial data science space. A strong global academic alliance provides a quality stream of SAS-literate graduates, and the educational offering of SAS is among the best in the industry. Still, almost all organizations focused on SAS face increasing challenges and complexity due to the proliferation of new programming platforms and paradigms, both in the industry and in academia. Inexpensive online learning options that mostly use alternative languages are already an established supplement, even an alternative, to academic or corporate training. They are one of the critical drivers of the democratization of data science. Publicly available SAS online courses remain mainly those developed by SAS, although the courseware developed by members of the SAS® Academic Alliance is mostly unavailable on leading MOOC platforms. In this session, we lay out a proactive career and professional development strategy for SAS professionals and organizations in this highly competitive environment. We look at MOOC platforms such as Coursera, Udacity, DataCamp, edX, and Udemy. We look at the success of the Kaggle concept (and some Kaggle-inspired failures). We propose an approach centered around SAS that leverages industry-focused communities such as Project Data Sphere, combined with SAS® Education, SAS Communities, SAS Academic Alliance, and popular MOOC platforms. Finally, we include lessons learned as the industry sponsor of a new SAS academic program.

INTRODUCTION

"Data science is a vague term, treat it accordingly" (Anonymous)

The massive proliferation of open source data science tools and libraries has radically changed the dynamics around SAS. On the one hand, SAS is no longer the language of initiation for data scientists (or data-centric scientists and domain experts). On the other hand, SAS became the only proprietary analytical platform that matters. Almost all formerly formidable competitors suffered a steep decline, with very few niche exceptions (such as Mathematica).

During the last few years, the demand for data scientists increased dramatically but so did the supply of pseudo data scientists. The feeding frenzy led to the proliferation of education and training offerings that unfortunately include many of dubious value, and the number of offerings is growing exponentially. Of particular concern is the spread of commercial training offerings that act as de-facto undercover recruiting boot camps where candidates essentially pay to be pre-screened and profiled by potential employers.

Most papers about learning in the data science context are concerned with employability or recruiting or the perceived shortage of talent. The focus of this paper is on retention, employee satisfaction, and the ROI from the existing talent pool.

Finally, this paper embraces open source while remaining unapologetically pro-SAS. We hold it self-evident that the field of data science greatly benefits from a strong commercial industry focused leader such as SAS. The often-cacophonous open source approach with its inherent lack of strategic consensus cannot be the only driver of innovation, strategy, and vision in such a rapidly changing field.

THE PIVOTAL ROLE OF THE ANALYTICAL CENTER OF EXCELLENCE

SAS defines the Analytical Center of Excellence (ACoE) as “An internal strategic team of experts with a specific focus on facilitating and promoting the use of analytics to achieve business objectives across the enterprise.” Like any center of excellence, the ACoE is a shared services unit whose purpose is to coordinate complex activities around newly introduced technologies.

Primary functions of the ACoE are:

1. Culture-driven leadership
2. Unification of analytical assets
3. Centralized learning and certification of resources for departmental deployment
4. Development and enforcement of standards, methodologies, governance (quality management)
5. Technology evaluation and introduction
6. Recruiting and talent management
7. Providing user communities with access to training
8. Common workspace in which methodologies, tools, models, and techniques are shared (knowledge management)

The existence of such a unit is a fundamental building block for a successful analytical learning organization. In the ideal case, the entire analytical workforce (data science & data engineering professionals) should be affiliated with the ACoE.

In case it is not possible to have an ACoE as a separate organization it is imperative to at least establish an ACoE as a virtual organization with its quality management system (QMS) and with overall control over skills and competencies.

There are two main objections to the ACoE concept. In large organizations, it is the opinion that analytics is no longer a new technology, that it is already something in the organizational DNA, and that the interaction can now be safely left to business and IT to negotiate without an intermediary. In a smaller organization, the objection is primarily that any CoE adds an additional organizational dimension. In the case of smaller organizations, one can build a valid argument. In larger organizations, it is a recipe for disaster. Considering the speed of technological change and the incoming (ongoing) AI hype cycle, the role of an ACoE as project office, knowledge management office, technology evaluation office, quality management office, but, most of all, as the organizational learning hub, remains indispensable.

RECRUITING AND TRAINING IN A UNIFIED MULTI-VENDOR ENVIRONMENT

It is a fact of life that SAS is no longer a part of the initiation ritual for most data scientists. During the last decade most people who come out of academia and claim “data science” skills and experience have never touched SAS but have a lot of experience on R or Python, and this trend will only continue (in absolute numbers). Nowadays, recruiting recent graduates or experienced data scientists and data engineers who will be working with SAS but have limited prior exposure to SAS skills should be perfectly acceptable, provided they have a mastery of open source tools and relevant domain expertise. Anyone who can master enough skills and experience for a data science role can easily pick up a new programming language, and SAS is not a difficult language to learn. Data scientists and

data engineers are expected to be continually learning new things, from languages to methodologies, and, as an employer, you would be sending a bad signal to existing and potential employees by not engaging with new technologies and sticking with whatever is perceived as “legacy”. Especially in the field of data science: curiosity is valued because it is part of the job to look at and question things (even programming languages).

With this in mind, the organization should be equipped to provide regular SAS-for-R or SAS-for-Python classes or boot-camps and should encourage utilization and unification of all analytical assets while preserving SAS as the hub holding the organization and analytical assets together. For this to take hold, leadership must introduce SAS Viya to the organization as soon as possible and take advantage of the fact that SAS is genuinely interested in helping organizations adopt Viya.

DEALING WITH THE LEGACY SYNDROME

We are in an era when there is a considerable focus on working with big data and data science, and the rising interest in those subjects will benefit both SAS and open source complementary platforms for a long time to come. Still, there is an unfortunate aspect of working with a stable platform on a daily basis: most of the work around it is routinely operational or maintenance-driven; this does not help. Data engineers like to keep their skills up to date, while data scientists (especially those with domain expertise) want to experiment, build algorithms, and innovate. Leadership needs to keep them happy, or they will leave. A few relatively simple strategies seem to work:

1. Encourage and reward participation in SAS User Groups and submission of papers for the SAS Global Forum. Executives should lead by example.
2. Create a “Lab” within your ACoE and budget time for “R&D” activities (algorithm development, use case development, testing of new technologies). Document findings, write blogs and publish papers.
3. Sponsor a data science meetup in the organization, geography or industry, and make sure that it does not fizzle out.
4. Encourage the adoption and use of open source tools in the SAS environment as well as the use of SAS in experimental or “bleeding edge” projects.
5. Bring Viya into the organization as soon as possible and pick a low-hanging fruit as the first project. Centralized model management comes to mind.

WHERE TO GET TRAINING? WHERE ARE IMPORTANT CONVERSATIONS TAKING PLACE?

SAS Education and SAS Communities are still by far the best non-academic places for SAS training and SAS-related discussions. Unfortunately, SAS professionals tend to be underrepresented in generic data science blogs, online communities, and live meetups which leads to the lack of SAS-related talks and posts on these platforms (other than the usual “Should I learn Python, SAS, or R?”).

The best conversations by far take place around SAS Communities and blogs. Authors are responsive, and community members show genuine interest, and Chris Hemedinger from SAS does a fabulous job as a curator. SAS Communities are also the best entry point to open source projects on GitHub. However, SAS Communities are not ideal. It would be beneficial to have a more considerable open source development effort curated by SAS that attracts developers from customer and partner organization. It would also be helpful to start drawing more developers with a non-SAS programming background. Last, but not least, communities have the untapped potential to become the hub for co-innovation which brings back the topic of SAS-curated open source projects.

SAS Education still provides the best training for SAS. That also includes e-learning. Some e-learning courses are offered for free, and the number of free e-courses will hopefully increase. There are also excellent "How To" free online tutorials. The high-end training offering by SAS currently includes SAS Academy for Data Science, SAS Academy for Clinical Programming, and SAS Certified AI and Machine Learning Professional track.

There are two ways to play with SAS software for free: using the SAS University Edition and using the SAS Viya Programming Trial.

SAS University Edition is available to all individual learners and now includes the ability to write Python. SASPy is an interface to SAS that automatically generates SAS code from Python 3.x scripts. The SASPy module can be imported into JupyterLab for SAS University Edition. From there you can create a connection to the SAS server, and then use Python objects and syntax to access SAS data and your SAS software. Of particular interest to Python aficionados is the ability to exchange data using the pandas software library and view your SAS results in your Jupyter notebook. You can also see the SAS code that is automatically generated.

SAS Viya Programming Trial is the easiest way to get initial exposure to Viya. There's nothing to download; a web browser is all you need. SAS provides sample data and code demo scripts as well as access to videos, documentation, and communities. You get eight hours to explore the software and if you need more time, you can save your work locally, and start a new trial. Virtually, you can explore the system as much as you want.

The trial gives you access to several SAS software solutions running on SAS Viya, such as SAS Visual Data Mining and Machine Learning and is highly recommended for all SAS and non-SAS data scientists and developers.

THE ROLE OF MOOC AND BOOTCAMPS

Massive Open Online Courses (MOOCs) are free online courses available for anyone to enroll. Leading aggregators include Coursera, edX, Udacity. The selection of SAS-related courses on MOOC platforms is relatively limited. The only SAS-specific MOOC offering is a SAS Programming Specialization series on Coursera (offered by SAS).

According to Vincent Granville from Data Science Central, there are three categories of non-academic data science programs: cheap, expensive, and respected. This classification implicitly refers to external offerings primarily targeted to individual learners. To these three we add the in-house learning program and boot camp.

Cheap courses are usually based on a monthly subscription model. It typically costs less than \$30 per month and gives access to a vast array of online classes. An example is DataCamp. Udemy seems to be the "people's platform." On this day, there were 199,806 students enrolled in the "Python for Data Science and Machine Learning Bootcamp" course on Udemy. While useful for individual learners, these platforms have limited use for enterprise learning.

Expensive data science and data engineering boot camps are intensive hands-on programs that frequently include tutoring and one-on-one mentorship. Some of the better-known boot camp providers are NYC Data Science Academy, Dataquest, General Assembly, Metis, Data Science Dojo, and DataCamp. Unless one wants to create a private boot camp, SAS Education is often the way to go. Ideal boot camp should be run by the ACoE as a core function, include the concept of mentorship, and should be jointly developed in cooperation with an experienced training provider and with SAS education.

Respected course offerings can be found on Coursera, edX, and Udacity. Recommended offerings include Data Science Specialization by Johns Hopkins University (Coursera),

Applied Data Science with Python Specialization by University of Michigan (Coursera), Machine Learning with Python by IBM (edX), and Advanced Machine Learning Specialization by Higher School of Economics (Coursera).

Before incorporating any of these offerings into the learning strategy, it is advisable to look into total cost and benefits in comparison with a structured offering, such as the SAS Academy for Data Science, which may turn out to be the most effective way to go, even cost wise.

In House. The best way to learn is to teach others. If your organization does data science for a living, you will gradually have enough material and experience to produce your own “Data Science Using SAS” course or boot camp. A third party may run the boot camp (e.g., SAS education or someone like Metis) but the organizational ownership should remain in the ACoE. If your courseware can be packaged as an online offering (e.g., on your in-house learning management system), you may want to consider sharing it externally on one of the MOOC platforms.

KAGGLE, “KAGGLE FOR SAS,” PROJECT DATASPHERE

Kaggle is a popular online community of data scientists, owned by Google. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions. It is important to know that SAS University Edition can be used to access Kaggle datasets and that SAS was historically used to develop code for Kaggle competitions.

Inspired by Kaggle, vendors such as IBM tried to build their competition platforms, but they were not able to attract enough interest. Advocating for a “Kaggle for SAS” is therefore probably a bad idea. On the other hand, open source projects that use SAS and are curated by SAS on GitHub would be an excellent way to co-innovate with SAS if they would attract more participants from customer and partner companies.

An interesting publicly accessible platform is Project Data Sphere, a free digital “library-laboratory” that provides one place where the research community can share, integrate and analyze historical, patient-level data from academic and industry phase III cancer clinical trials. Anyone interested in cancer research can apply to become a user and get access to data that can be analyzed using SAS tools that are also freely available on the platform.

INNOVATION IN THE SAS LEARNING ECOSYSTEM

Learning ecosystems around commercial offerings are understandably less entrepreneurial and innovative than around open source. The community around the commercial offerings is used to the vendor taking the lead and setting the strategy; after all, that’s why we pay them. As of recently however, there is an interesting player in the SAS educational ecosystem that defies this rule.

SAS Sensei or SASSENSEI (<https://sasensei.com>) is a training tool developed by Allan Bowe from the UK built according to the principles of gamification of learning. It covers a range of topics including SAS Programming, Administration, Viya, Risk Dimensions, DI Studio, AF / SCL, Stored Process Web Apps, and more. According to the authors, it is a quiz site that leverages the concepts of question-based learning and community validated content. It has been created to provide a knowledge base by which players can:

- Learn (using the flashcards)
- Test (via questions and quizzes)
- Understand (by writing and reviewing intelligible content)

By answering questions and passing quizzes, and by submitting quality content, one will earn tokens and reputation that can be used to unlock the later levels.

This approach has the potential to reach users and generate user engagement in a way traditional offerings have failed.

ALIGNING WITH A SAS ACADEMIC PROGRAM

Established “Master in Analytics” programs at US universities that teach SAS are generally of good quality and most are very pragmatic and open to cooperating with the industry. Many students, especially international, enroll in these programs after a couple of years of work experience and graduate with some domain expertise in their portfolio. By aligning with such a program at a business-friendly college, one can create a steady pipeline of interns and potential recruits. If you can offer exciting internships, there is practically no downside and benefits are many. Our company cooperates with such a program.

Our company also teamed with SAS to create an academic program at a large university in Europe. Getting it off the ground was more challenging than expected. We discovered that it is a long-term process that requires at least a 3-year commitment between the university and the commercial sponsor and that on-campus recruiting is not a sufficiently strong glue between the program and you as the industry sponsor. A joint project that employs students (especially doctoral students) and brings funding to the university, or a structured internship program, will significantly accelerate the process.

CONCLUSION

The objective of each analytical organization is to provide value to the business. In this rapidly changing climate, learning management must be a critical leadership responsibility. Your organizational culture is best defined and most impacted by your internal learning culture, while your ability to offer learning and mentorship to business stakeholders helps build the fabric of a “joint venture” between IT and business, making and keeping analytics a part of the organizational DNA.

REFERENCES

Zeid, Aiman; Milley, Anne, “How to Establish an Analytical Center of Excellence to Maximize the Value from Your Data and Analytics Investment.” Proceedings of the SAS Global Forum 2010. Available at: <http://support.sas.com/resources/papers/proceedings10/173-2010.pdf>

Zeid, Aiman; “Best Practices for Analytics Center of Excellence.” Available at: https://www.sas.com/content/dam/SAS/en_ca/User%20Group%20Presentations/Ottawa%20User%20Group/BestPracticesforAnalytics.pdf

Galarnyk, Michael. “How to Build a Data Science Portfolio.” Jul 8, 2018. Available at: <https://towardsdatascience.com/how-to-build-a-data-science-portfolio-5f566517c79c>

Bowe, Allan. 2018. “Find Your SAS® Sensei,” Proceedings of the SAS Global Forum 2018, Denver, Co. Available at <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1792-2018.pdf>

RECOMMENDED READING

- SAS Viya: The Python Perspective (SAS Institute). 2017. Kevin D. Smith, Xiangxiang Meng
- SAS Viya: The R Perspective (SAS Institute). 2017. Yue Qi, Kevin D. Smith, Xiangxiang Meng

- Coding in Python with SAS University Edition (SAS Blog). 2018. Chris Hemedinger. Available at: <https://blogs.sas.com/content/sasdummy/2018/02/05/python-sas-university-edition/>
- Introducing SASPy: Use Python code to access SAS (SAS Blog). 2018. Chris Hemedinger and Jared Dean. Available at: <https://blogs.sas.com/content/sasdummy/2017/04/08/python-to-sas-saspy/>
- Business Intelligence Competency Centers: A Team Approach to Maximizing Competitive Advantage (Wiley and SAS Business Series Book 12) Jul 6, 2007. Gloria J. Miller, Stefanie V. Gerlach, Dagmar Brautigam
- Out in the Open with Analytics: Combining the Benefits of SAS and Open Source Analytics, by SAS Institute. Available at: https://www.sas.com/en_us/whitepapers/open-analytics-109311.html
- SAS Programming for R Users. Course Notes. Available on GitHub at https://github.com/sassoftware/sas-prog-for-r-users/blob/master/notes/LWSP4R_001.pdf

SEE ALSO

To enjoy analytics and data science, you need to have interesting data sets to play with. A large number of exciting datasets is available in the public domain. Rather than listing them here, I recommend searching Quora for “Where can I find large datasets open to the public.”

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jovan Marjanovic
 ProWerk Consulting
 703.342.7402
jovan.marjanovic@prowerk.com
www.prowerk.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.