The Futurum Group

Lab Insight

From Data to Decision: Increasing AI Productivity with SAS Viya

AUTHORS

Mitch Lewis
Research Analyst | The Futurum Group

COMMISSIONED BY

SAS

SEPTEMBER 2024

Productivity Study Overview

Organizations faced with a wide range of business challenges are increasingly turning to data-driven approaches to find new and innovative solutions. The emergence of machine learning and AI – and now generative AI – has enabled organizations to leverage their data to form valuable business decisions, create new products, and drive innovation. While data serves as a powerful resource for businesses to leverage, the process of transforming raw data into valuable information and insights is not without its challenges.

Organizations rely on teams of Data Scientists, Data Engineers, and MLOps Engineers to transform data into actionable insights. This is achieved through a process known as the data and AI life cycle, which spans from managing and preparing data, through creating and optimizing AI models, to deployment and monitoring. While the data and AI life cycle has become a crucial component of core business processes, data and AI teams face challenges of time, complexity, and resources at each step. Organizations capable of increasing productivity throughout this life cycle can achieve significant benefits including faster decision-making, lower costs, and greater innovation.

Key to navigating the data and AI life cycle are the tools and platforms used by data and AI practitioners. This study presents an evaluation of three data and AI platforms, with a focus on how they impact productivity. Productivity was evaluated by measuring the engineering time required to complete common data and AI tasks, along with additional evaluation of the complexity and skill requirements involved. The study involved a hands-on evaluation from multiple Futurum Group analysts throughout the data and AI life cycle in SAS® Viya® along with two alternative solutions. Notably, the study found that SAS Viya enabled significant productivity gains throughout the data and AI life cycle compared to competitive solutions.

Key findings of the impact of SAS Viya on productivity included:

4.6x
more productive on average throughout the end-to-end data and AI life cycle

No coding requirements throughout the entire life cycle

86%
of all tasks were determined to be achievable by a Business Analyst

The Data and AI Life Cycle

Although organizations have identified data as a key resource, raw data alone is not enough to drive business decisions. Instead, data must be processed alongside a company's contextual knowledge to extract and deploy the insights it contains. This is the role of the data and AI life cycle.

The data and AI life cycle is an ongoing, iterative process undertaken by teams of Data Scientists, Data Engineers, and MLOps Engineers that can broadly be broken into three core components: Manage Data, Develop Models, and Deploy Insights. The exact steps involved in the life cycle may vary between organizations and their specific data challenges; however, in general, these three components account for the main processes required of an organization to derive business insights from their raw data.

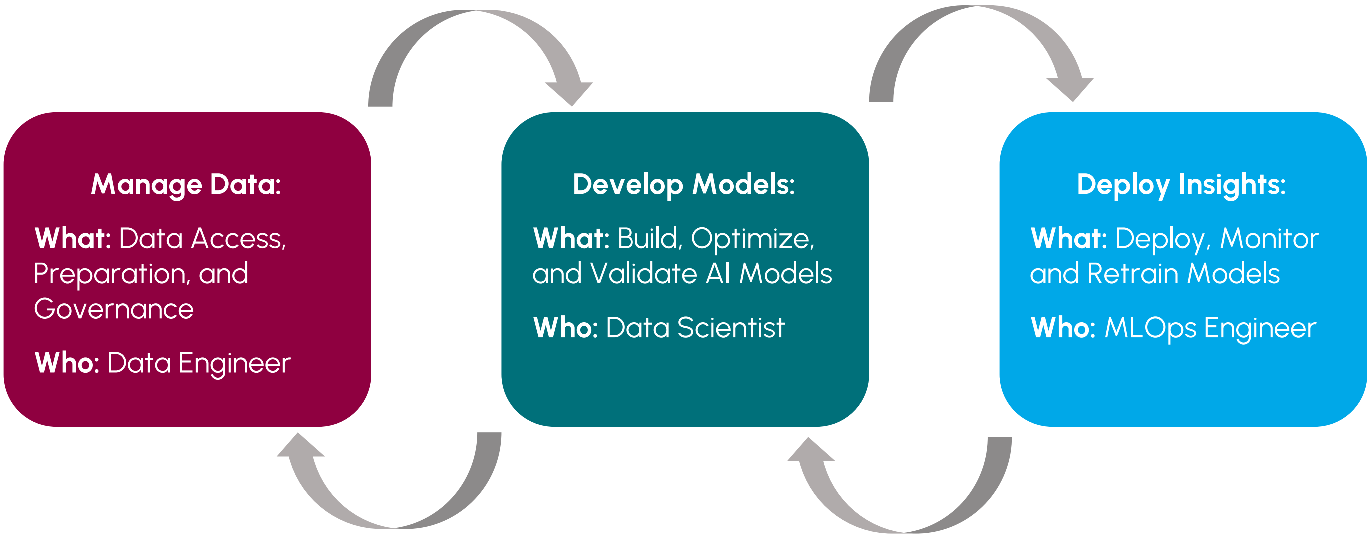


Figure 1: The Data and AI Life Cycle

As depicted in Figure 1, data flows forward through the life cycle and then backward as needed to iterate on a previous step. The first key step in the life cycle is managing data, which involves accessing data from various sources, ensuring data meets governance and data privacy constraints, and finally cleaning and preparing the data to meet the requirements of data modeling. This process is typically the responsibility of a Data Engineer. From there, a Data Scientist can utilize this data to develop models. This process typically involves data exploration and feature engineering, along with an iterative process of training, optimizing, and validating AI models, before registering a champion model to be deployed. At this stage of the life cycle, the champion model can be deployed by an MLOps Engineer, who is additionally responsible for monitoring the ongoing performance of the model, automating retraining to avoid model degradation, and building additional automation into processes where possible. This flow represents a general outline of the processes and roles involved in the data and AI life cycle; however, it should be noted that the specific tasks and stakeholders responsible for them may vary.

Increasingly, this life cycle plays a crucial role for modern organizations. By utilizing the data and AI life cycle, organizations are able to drive business decisions, gain new insights, develop new data products, and ultimately increase innovation.

Productivity in Data and AI Life Cycle

As key business decisions and processes continue to be influenced by the data and AI life cycle, the productivity of data and AI teams navigating this life cycle becomes increasingly important. The life cycle can be complex, involving significant time and resources, which can hinder overall productivity. Meanwhile, increasing productivity has the potential to provide faster business decisions, further enhance innovation, reduce costs, increase revenue, and create an overall competitive advantage.

To examine the productivity of the data and AI life cycle, it is important to understand the challenges commonly impacting data and AI teams and how these challenges can be overcome to increase productivity. Key challenges identified include the following:

- **Data**

  Data is the key element in the data and AI life cycle; however, it also commonly poses challenges that dramatically reduce productivity. Organizations often have several data sources, challenging them to find insights across all of their data. Additionally, data is often messy – with challenges including inconsistent structure, missing values, and incompatible data types. Typically, significant portions of the data and AI life cycle are spent cleaning and preparing data, rather than modeling or utilizing it. Data may additionally pose privacy or governance issues, as any private or sensitive data must be identified and properly obscured.

  To address these challenges, organizations should seek platforms that not only offer support for all of their required data sources, but also provide tools that assist with data preparation and data governance concerns.

- **Talent**

  The complexity found at each stage of the data and AI life cycle has typically limited participation to highly skilled, expert resources. These expert resources, including Data Scientists, Data Engineers, and MLOps Engineers, are rare and can prove challenging for organizations to find and obtain. Employing expert data practitioners commonly involves significant labor costs, and organizations should seek to maximize their productivity. Organizations capable of successfully upskilling workers or utilizing less technical resources may be able to mitigate these challenges.

An additional talent challenge can be ensuring efficient collaboration. The data and AI life cycle not only requires significant collaboration between the technical experts responsible for each phase, but it may also require collaboration with additional non-technical or business-focused users as well. While advanced technical skills are typically required to accomplish the data and AI life cycle, collaboration among additional stakeholders can prove extremely beneficial when applying the life cycle to business-specific challenges. Where such collaboration is required, organizations should select tools that both enable efficient collaboration and provide an intuitive environment for non-technical resources.

- **Tools and Technology**

  While various tools and technology can significantly assist in navigating the data and AI life cycle, they can also present several challenges. Tools utilized across the data and AI life cycle are often considered complex, requiring a significant learning curve and advanced programming knowledge. As technology evolves, interoperability between new and existing tools may also pose a challenge. This is especially notable with the rapid and ongoing advancements in the field of AI seen in recent years. This challenges organizations to find solutions that are technically capable, interoperable, and easy to use.

- **Infrastructure**

  The complexity of the data and AI life cycle additionally complicates the required infrastructure. Organizations are challenged both with procuring infrastructure capable of supporting demanding AI and analytics workloads, as well as taking on the time and complexity of managing such infrastructure. Significant time and effort spent managing infrastructure may ultimately impact time spent navigating the actual data and AI life cycle.

- **Time**

  One of the most important factors impacting productivity is time. When considering the data and AI life cycle, there are both human and computation time requirements that impact productivity. Computation time often comes under consideration through the lens of a solution's performance capabilities, while the human time component is often intertwined with many of the productivity considerations previously mentioned.

With these key areas identified, data and AI teams can focus on ways to mitigate such challenges, and ultimately increase productivity. Often, this may require leveraging more innovative tools and technology; however, as previously discussed, such tools should involve a low learning curve and interoperability with additional technologies and data sources. Among new tools available, cloud services have gained popularity as they additionally help address infrastructure challenges and provide a centralized platform for collaboration. The impact of time and labor can also be reduced by implementing tools that provide significant automation capabilities and reduce the experience level and technical expertise requirements.

Data and AI Tools and Approaches

Key to solving productivity challenges throughout the data and AI life cycle are the tools and technology leveraged by data and AI teams. As data analytics, machine learning, and AI have become increasingly popular approaches to solving business challenges, the market for tools supporting these processes has additionally evolved. In general, there are two distinct approaches for organizations to take: commercial and non-commercial.

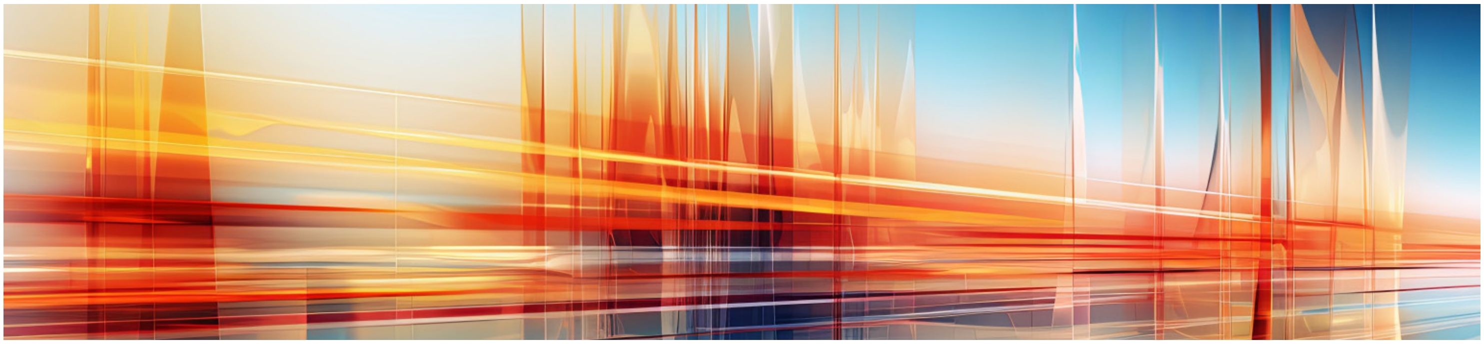
A non-commercial approach combines several open source tools together to address each phase of the data and AI life cycle. The exact open source tools and packages selected may vary depending on the preference and specific requirements of an organization; however, there are a large number of open source technologies available with capabilities to complete most common data and AI tasks. While this type of non-commercial, open source approach provides significant flexibility, it can result in a highly complex solution. This approach is heavily reliant on programming capabilities, with additional knowledge required for external packages and tools. This can be time consuming and require a team with significant experience utilizing such tools.

Along with flexibility, the other main benefit of selecting a non-commercial approach is the economic benefit associated with avoiding any licensing or subscription costs found in commercial options. While this cost avoidance is a major incentive for many to select a non-commercial approach, organizations should also be aware that constructing their own open source-based platform can add significant complexity. This approach leaves organizations responsible for major considerations, such as management of both infrastructure and various software dependencies and its associated risks. This added layer of management, without built-in support offered from a commercial platform, can add time, complexity, and associated risks that may ultimately limit productivity and add costs.

Alternatively, organizations can leverage one or more of several commercially available data and AI platforms. The commercial data and AI platform market has evolved with growing demand from data-driven organizations, often hindered by many of the previously defined challenges. Modern commercial platforms are primarily cloud-based, offering a convenient centralized platform to combine data sources and enable collaboration, while simultaneously removing infrastructure management requirements. Such commercial data and AI platforms additionally package built-in features and automation capabilities to assist organizations in effectively navigating the data and AI life cycle.

As the market for commercially available data and AI platforms has evolved, various platforms have emerged, with different fundamental approaches and capabilities. While all commercial data and AI platforms should be capable of facilitating the core tasks within the data and AI life cycle, different platforms will provide varying levels of functionality and complexity.

The goal of this study was to evaluate the overall productivity differences found between SAS Viya, a competitive commercial data and AI cloud platform, and a non-commercial platform, based on open source technologies.



About the Study

The Futurum Group conducted an in-depth analysis of three distinct data and AI platforms to measure their impact on productivity throughout the data and AI life cycle:

- SAS Viya – a commercially available cloud native data and AI platform
- A competitive commercial data and AI cloud platform
- A non-commercial platform, based on open source technologies

In addition, MLflow, an open source MLOps platform, was leveraged alongside both the commercial and non-commercial platforms to enhance their MLOps capabilities and create a fair comparison with SAS Viya.

To compare productivity, each platform was used to complete a single pass through the end-to-end data and AI life cycle. To create an even comparison, each platform was used to solve the same business challenge. For this testing, the goal was to complete a customer churn predictive analysis with simulated banking customer data. While this evaluation specifically utilized banking data, a customer churn predictive analysis was chosen as it is a common data and AI use case that is applicable to a wide range of industries. Testing involved two core datasets, both formatted as csv files.

- Customer Data
    - 10,095 rows
    - 39 columns
- Bank Account Data
    - 10,088 rows
    - 19 columns

This data was selected to provide a realistic sample suitable for completing all tasks throughout the data and AI life cycle. As this study was not focused on evaluating system performance, a relatively small sampling was utilized.

To simulate a realistic approach to the data and AI life cycle, testing was completed by four distinct analysts, each assigned to specific personas responsible for a portion of the end-to-end life cycle. Analysts were matched to their specific persona based on their technical expertise and prior knowledge of each area. To create a fair evaluation of each platform, no analyst involved had significant prior experience using any of the three platforms evaluated. This approach allowed each phase of the data and AI life cycle to be individually evaluated through the persona of a well-aligned professional role. It also allowed for evaluation of collaboration among distinct members of a data and AI team.

Testing personas included:

- **Data Engineer** - Responsible for the core tasks aligned with the Manage Data phase of the data and AI life cycle
- **Data Scientist** - Responsible for the core tasks aligned with the Develop Models phase of the data and AI life cycle
- **MLOps Engineer** - Responsible for the core tasks aligned with the Deploy Insights phase of the data and AI life cycle
- **Business Analyst** - Additional persona responsible for testing all tasks in the end-to-end data and AI life cycle to evaluate achievability by a non-technical resource.

To achieve the analysis, each persona completed all required tasks in each environment and recorded several metrics impacting productivity, including the time, complexity, coding requirements, and expertise required. The parameters of each metric were defined as follows:

- **Time** - Time was recorded as a measurement of the "Engineering Time" required to complete each task. This measurement isolated the time required of a human to complete a given task and was not a measurement of computational time or system performance. Performance testing was considered outside the scope of this analysis; however, additional performance-focused testing of data and AI platforms previously conducted by The Futurum Group can be found here.

  To avoid the confusion between time as a measurement of human effort and computational performance, the time-based comparisons within this study have been recorded as a measurement of productivity, rather than with generic terms such as "faster".

- **Complexity** - While complexity is a subjective measurement, it can be a crucial factor impacting productivity. To capture the complexity of completing a given task, The Futurum Group analysts established a common framework to assign each task a complexity score of "High", "Medium", or "Low" The complexity score of each task was determined by the individual persona completing the task, with an assumption of a baseline set of skills and experience that would be commonly expected from a Data Engineer, Data Scientist, or MLOps Engineer, respectively. Factors considered within the complexity score included:

  - The learning curve required to complete the task, as measured by the amount of time and learning required to comfortably complete the task in each environment.
  - Technical skills or knowledge that were considered to be beyond the common expectations of each persona. This included tasks that were outside the scope of a typical Data Engineer, Data Scientist, or MLOps role, as well as where additional tools or software was required. Utilization of additional tools was found to add complexity in several ways, including creating dependencies or compatibility issues as well as time spent reading documentation and troubleshooting.
  - Significant amounts of coding required to complete a given task. Coding requirements were considered minimal for tasks that could be completed within a few lines of code. Significant coding requirements included areas in which the coding required was well beyond a few lines of code or was dependent on the use of advanced programming libraries.

- Requirements of additional packages, installation, or system management involved in the completion of a given task. These steps add complexity by involving additional time and expertise not directly contributing to the completion of the data and AI life cycle.

- **Coding Requirements** - A binary "Yes" or "No" was recorded for each task, dependent on if programming was required to successfully complete the task.

- **Achievable by Business Analyst** - Each task was graded by the Business Analyst persona as to whether it was reasonably achievable by a Business Analyst or other non-technical resource with either a "Yes" or "No" This was determined by evaluating the technical skills and knowledge required of a task in relation to those of a typical Business Analyst who is not considered an expert in data engineering, data science, or MLOps.

Key Highlights

"Testing showed that an end-to-end data and AI life cycle can be achieved with more than 4x greater productivity in SAS Viya than in competitive solutions. The productivity gains found in SAS Viya not only ease the burden placed on data and AI teams, but they can lead to significant cost savings."– *Mitch Lewis, The Futurum Group*

Evaluation of the data and AI life cycle in each environment found that SAS Viya enabled several significant productivity gains compared to both the competitive commercial platform and the non-commercial environment. Several findings from this testing stand out as key factors in increasing productivity.

Notably, SAS Viya was found to enable faster navigation of the data and AI life cycle. Testing found that achieving the life cycle was up to 4.25x more productive than in the commercial platform, and 5x more productive than in the non-commercial environment, for an average of 4.6x greater productivity. These findings show that SAS Viya enables data and AI teams to achieve the end-to-end data and AI life cycle in less than a quarter of the time required in other competitive environments. Along with enabling more efficient navigation of the end-to-end life cycle, SAS Viya was also found to provide a time advantage during each of the three individual phases of the data and AI life cycle.

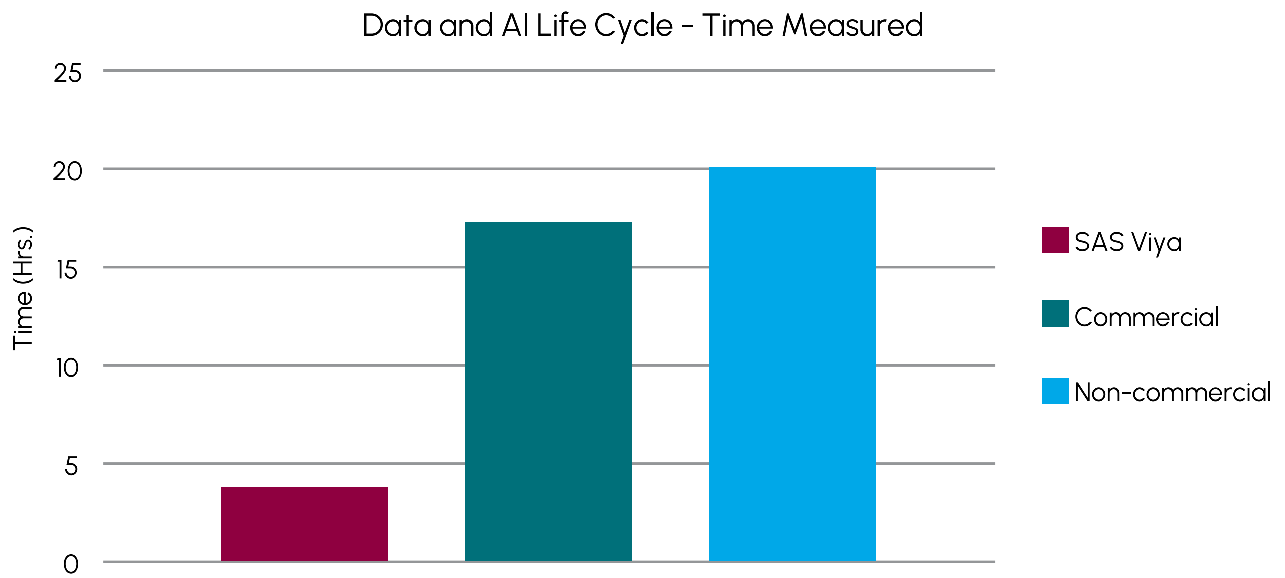


Figure 2: Data and AI Life Cycle Time Measured

When evaluating complexity, 96% of all tasks completed in SAS Viya were rated as "Low Complexity", with the remaining 4% considered "Medium Complexity". No tasks were considered "High Complexity" in SAS Viya. Comparatively, the commercial platform was found to achieve "Low Complexity" in 50% of tasks and "Medium Complexity" in 33% of tasks. In the non-commercial environment, 30% of tasks were considered "Low Complexity" and 53% were considered "Medium Complexity". In both competitive environments, 17% of all tasks were considered to be "High Complexity".

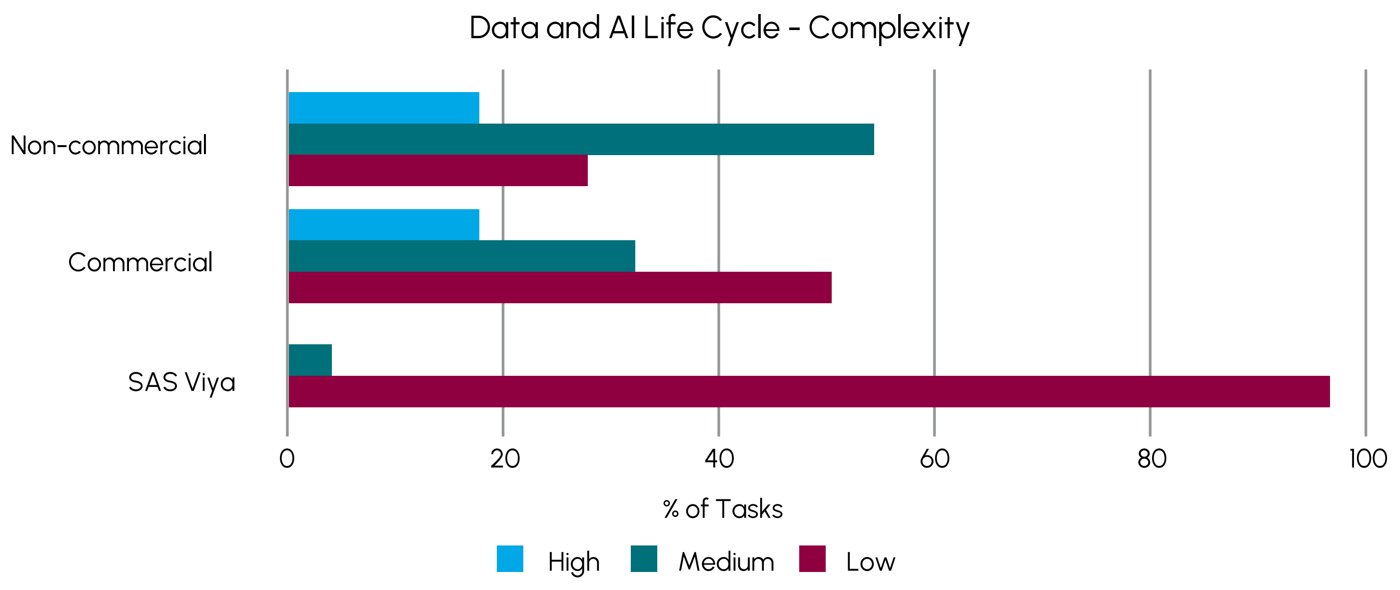


Figure 3: Data and AI Life Cycle Complexity

SAS Viya was also found to achieve the entire data and AI life cycle without any coding requirements. While it was noted that SAS Viya provided additional flexibility to complete tasks with SAS, Python, or R code, all tasks tested could be sufficiently completed utilizing built-in automation or no-code operations in the SAS Viya GUI. Alternatively, both competitive platforms evaluated were found to require significant coding to achieve the life cycle. The competitive commercial platform was found to require coding for 50% of all tasks tested, with some built-in features capable of reducing coding requirements for specific tasks. The non-commercial platform was found to be a primarily code-focused approach, with 77% of all tasks requiring coding to complete.

Evaluation by the Business Analyst persona revealed additional productivity advantages for SAS Viya, with findings showing that a greater percentage of the overall life cycle could be completed by a Business Analyst resource. In SAS Viya, 86% of the end-to-end life cycle was found to be achievable by a Business Analyst, compared to 56% in the alternative commercial environment and 47% in the non-commercial environment.

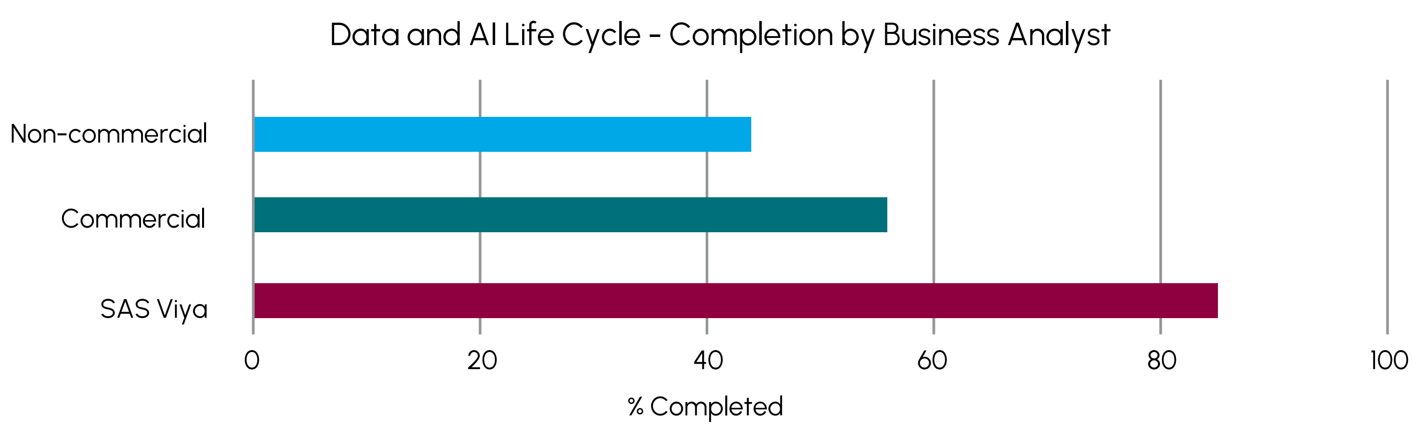


Figure 4: Data and AI Life Cycle Completion by Business Analyst

Data Engineer Findings

> ***"Data engineering tasks were achieved with 16x greater productivity in SAS Viya than in competitive platforms. By reducing the time managing and preparing data, organizations can instead focus on utilizing their data to innovate processes and drive business decisions." –*** *Russ Fellows, Data Engineer*

The responsibility of the Data Engineer was to complete the Manage Data portion of the end-to-end data and AI life cycle. This persona was tasked with evaluating and preparing the raw data into an analytical base table that could then be utilized by the Data Scientist.

The specific tasks completed by the Data Engineer in each environment were:

- **Data Upload** - This process involved the ingestion of the two CSV data files used for the churn analysis into the environment under evaluation.
- **Data Profiling** - Uploaded data was reviewed to examine the format, data types, and metadata information.
- **Data Sensitivity Analysis and PII Identification** - Data was reviewed to find and flag sensitive or personally identifiable data.
- **Data Quality Analysis** - Data was visualized and investigated to find any data quality issues, such as missing, duplicate, or outlier values.
- **ETL** - Data was cleaned and transformed for use by the Data Scientist with an ETL process including the following steps:
  - Deduplication
  - Joining tables
  - Data masking
  - Value binning
  - Save final table

In total, the Data Engineering workflow in SAS Viya was found to be achievable in approximately 18 minutes, with zero coding required, and 100% achievability by the Business Analyst persona. Alternatively, the same process required approximately 5 hours in both competitive environments with 40% completion by the Business Analyst persona. Overall, SAS Viya was found to enable 16x more productive completion of core Data Engineering tasks than either competitive environment, as well as enable greater utilization of business analysts.

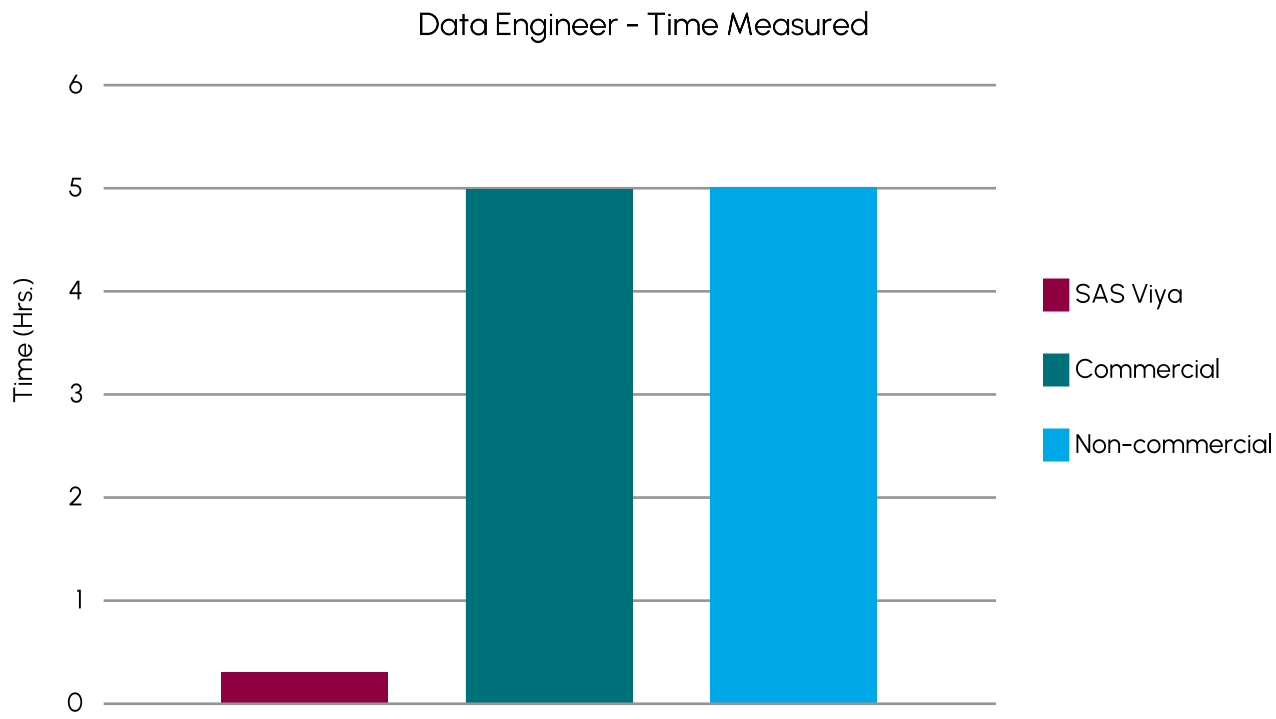


Figure 5: Data Engineer Time Measured

The productivity advantages found throughout the Data Engineer workflow were a result of several built-in automations within SAS Viya, as well as leveraging no-code processes that were found to be simple and effective. Notably, SAS Viya provided automated data scanning functionality that was utilized to quickly examine data and identify sensitive data. Neither competitive platform included this type of automation and instead require a manual approach. In addition, it was noted that this manual process introduced greater potential to miss sensitive or private data. The datasets used in this testing were relatively small and contained sensitive data that was easily distinguishable; however, with larger and more complex datasets, these challenges may become increasingly complex and time consuming for Data Engineers.

Additionally, SAS Viya included several automated data visualizations that enabled efficient analysis of data quality. Alternatively, both competitive platforms required time spent creating comparable visualizations. The commercial platform did enable visualizations to be created in a low-code dashboard environment, which may be more familiar to a Business Analyst; however, this process was not found to save time compared to a coding-based approach.

The largest differentiator found between SAS Viya and the two competitive platforms was in creating the ETL pipeline. This process was achievable in SAS Viya with a no-code flow builder, in which all steps were easily completed through a drag and drop interface. SAS Viya additionally automated some data cleaning processes, such as data encoding, which further increased the productivity gains. In total, the ETL process in SAS Viya was found to be achievable in approximately 6 minutes. In both the commercial and non-commercial platforms, the same ETL process required significant coding to clean and transform the data into a table prepared for a Data Scientist. This coding process required approximately 4 hours, along with extensive knowledge of data manipulation and data cleaning best practices.

Data Scientist Findings

"SAS Viya's built-in visualization tools reduced time spent on data exploration by 69x, freeing up data scientists to focus on core data modeling tasks and quickly achieving new insights." – *Frederic Van Haren, Data Scientist*

The second portion of the data and AI life cycle evaluation was undertaken by the Data Scientist persona, corresponding to the Develop Models portion of the data and AI life cycle. The responsibility of the Data Scientist was to explore the dataset prepared by the Data Engineer and utilize it to build a model to predict customer churn.

The specific tasks completed by the Data Scientist in each environment were:

- **Visual Exploration and Insights** - Creation of data visualizations to explore the attributes and trends contained in the dataset. This task was further separated into two distinct sections:
  - Visual Exploration and Insight Discovery
  - Visual Exploration - Augmented Analytics
- **Outlier Detection** - Detection of outlier values that may impact the accuracy of the models.
- **Quick Model Prototyping** - Development of a quick proof of concept model.
- **Model Building** - Creation of various machine learning models to predict customer churn.
- **Model Competitions** - Comparison between created models.
- **Model Tuning** - Hyperparameter tuning.
- **Explainability** - Generation of plots and visualizations to explain models' behavior.
- **Bias Detection** - Detection and visualization of bias within model variables.
- **Model Reports** - Generation of reports detailing models for regulatory and audit purposes.
- **Auto ML** - Experimentation of creating an automated machine learning model with AutoML capabilities.
- **Pipeline Competition** - Creation and comparison of different machine learning pipelines.
- **Model Registration** - Registering a selected champion model in a common repository to be made available to MLOps for model deployment.
- **Project Insight Reports** - Generation of reports detailing insights from the ML model and additional information about the project.
- **Sharing Projects** - Sharing of the data science project to make it available to additional team members.

All Data Scientist tasks were completed in SAS Viya in approximately 3 hours. Testing found that all tasks could be achieved without any coding requirements, and 80% of tasks were determined to be achievable by the Business Analyst persona. It should be noted that the remaining 20% of tasks could be completed without requiring advanced technical skills. The processes for these tasks were found to be intuitive; however, they required additional background knowledge of data science, beyond that of a typical Business Analyst.

In the commercial platform, the Data Scientist workflow required 3x as much time, with completion in approximately 9 hours. The non-commercial platform required 4x more time, with completion in approximately 12 hours. Both competitive platforms were additionally found to require greater expertise to complete, with 53% and 33% of tasks determined to be achievable by the Business Analyst in the commercial and non-commercial platforms, respectively. In these platforms, tasks were deemed unachievable by a Business Analyst due to a combination of both data science knowledge and technical skill requirements.

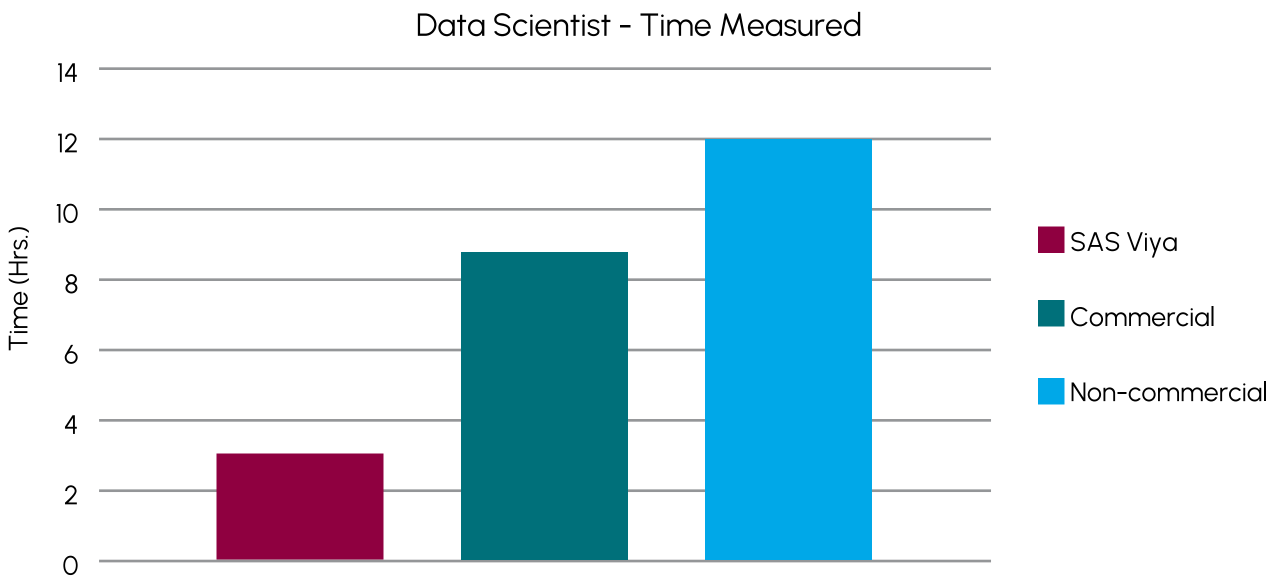


Figure 6: Data Scientist Time Measured

Productivity in SAS Viya was found to be greatly enhanced by built-in visualizations and automation, simplifying data exploration and reporting processes. SAS Viya additionally provided pre-built model templates that reduced time building models and simplified the hyperparameter tuning process.

In both competitive platforms, the Data Scientist workflow was found to be less automated and required significant coding. Significant time was spent on data exploration and manually creating data visualizations compared to those automated by SAS Viya. The model-building process was additionally found to be more time-consuming, requiring packages such as scikit-learn for model building and tuning and MLflow for tracking model development and comparing model performance. While these packages may be familiar to experienced Data Scientists, they were considered to limit the productivity of less experienced resources and add to the complexity and learning curve of the full workflow.

The non-commercial platform required additional packages to supplement features found in both commercial platforms, including auto-sklearn, SHAP, and LIME. The inclusion of several packages in a self-managed environment led to dependency issues, requiring additional time and effort of the Data Scientist persona.

MLOps Engineer Findings

"In SAS Viya, monitoring model performance was achieved with 8x more productivity than in competitive solutions, and retraining models was found to be 7x more productive. This efficiency provides MLOps teams with agility to quickly respond to any changes in model performance, ensuring high quality models are maintained." – *Brian Martin, MLOps Engineer*

The third step in the data and AI life cycle is Deploy Insights, which corresponds with the role of the MLOps Engineer persona. After completion of the Data Scientist tasks, the role of the MLOps Engineer begins, deploying the model and maintaining its performance over time.

The specific tasks completed by the MLOps Engineer in each environment were:

- **View Registered Models** - Find and view the models registered by the Data Scientist.
- **Generation of Model Deployment Files** - Creation of model artifacts, metadata, and associated code.
- **Model Comparison** - Review and visualize metrics of registered models.
- **Model Scoring** - Scoring of models with additional data before deployment.
- **Model Deployment** - Deployment of the chosen model into the production environment.
- **Monitor Model Performance** - Ongoing monitoring to ensure model performance. For this evaluation, additional datasets representing quarterly bank customer data were utilized.
- **Model Retraining** - Additional model training with updated data, as deemed necessary during performance monitoring.
- **Versioning** - Creation of a model version history for rollback purposes as well as to create an audit trail for compliance purposes.
- **Alerting** - Creation of automated alerts based on user-defined performance thresholds.
- **Scheduling** - Creation and scheduling of a workflow to run the model and score it against new data.

As with the Data Engineering and Data Scientist tasks, SAS Viya was found to enable greater productivity throughout the MLOps tasks. All MLOps tasks were found to be completed in approximately 40 minutes, 4.5x more productive than the competitive platforms, which were both found to require approximately 3 hours. With SAS Viya, it was found that 90% of the MLOps tasks evaluated could be completed by the Business Analyst, compared to 70% in both competitive platforms.

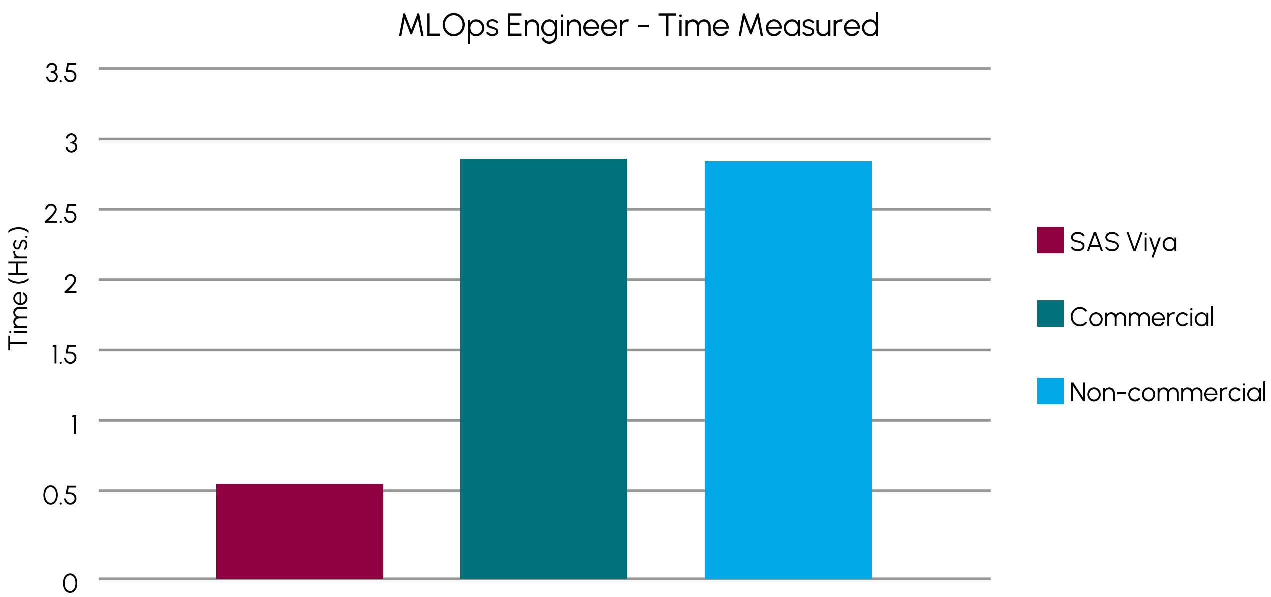


Figure 7: MLOps Engineer Time Measured

The productivity of the MLOps Engineer was found to be greatly enhanced by automated model performance reporting available in SAS Viya. This functionality enabled model performance monitoring to be achieved in approximately 15 minutes. Since the model reports and visualizations were generated automatically, the majority of this time involved analyzing and understanding the report. In both competitive platforms, performance monitoring was found to account for a significant portion of the total MLOps workflow, requiring approximately 2 hours in both environments. This process involved greater complexity in calculating model performance metrics and creating equivalent visualizations, along with similar requirements to analyze and understand the results. In both competitive platforms, MLflow was utilized to enable the MLOps workflow.

It should be noted that the competitive commercial platform did include an automated approach to performance monitoring. After testing this approach, however, The Futurum Group analysts determined that it did not provide effective results, and they reverted to a more manual approach, similar to that achieved in the non-commercial version. The Futurum Group believes that with additional enhancement to this functionality, model performance monitoring could be achieved much quicker. Given the assumption that future enhancements to this functionality will provide reporting similar to what was found in SAS Viya, The Futurum Group estimates that model performance reporting could be achieved in the commercial platform in approximately 30 minutes. This estimate assumes the time spent reviewing reports is similar to SAS Viya, while accounting for some additional complexity that was found in configuring the feature.

Model retraining was additionally found to be greatly simplified in SAS Viya, providing productivity gains for the MLOps Engineer. **In SAS Viya, model retraining can be quickly achieved with a point-and-click process. Alternatively, model retraining in both competitive platforms involved greater time and effort, requiring detailed knowledge of machine learning and associated programming libraries**, such as scikit-learn and the MLflow platform, which was used to support MLOps tasks.

Business Analyst Findings

> ***"By reducing the skillset required to complete key tasks within the data and AI life cycle, not only does SAS Viya unlock greater contributions from Business Analysts, it frees up time for expert resources to focus on new innovations." –*** *Mitch Lewis, Business Analyst*

The final persona utilized to evaluate the three data and AI platforms was that of a Business Analyst. The role of the Business Analyst was to attempt all tasks in the end-to-end data and AI life cycle to determine if they were realistically achievable by a non-technical resource. The Business Analyst persona represents additional roles often associated with data and AI teams that are not distinct experts in Data Engineering, Data Science, or MLOps. While this study chose the term "Business Analyst", the capabilities of this persona may apply to roles with a wide range of titles. The Business Analyst persona was established as an individual familiar with data collection, analysis, and visualization primarily through the use of low-code data visualization and dashboard tools. The Business Analyst was also assumed to have familiarity with SQL and some basic programming knowledge. The Business Analyst persona was assumed to lack in-depth knowledge or experience with core Data Engineering, Data Science, or MLOps skills. This includes both technical skills, such as using advanced programming libraries, as well as background knowledge, theory, and best practices of each discipline.

The findings of the Business Analyst were as follows:

- 86% of tasks were achievable in SAS Viya.
- 56% of tasks were achievable in the commercial platform.
- 46% of tasks were achievable in the non-commercial environment.

Key to the high completion rate of tasks in SAS Viya was the ability to achieve all tasks with zero coding required. SAS Viya was found to provide an intuitive, low-code environment that could realistically be utilized by a Business Analyst to achieve the majority of tasks in the end-to-end data and AI life cycle. It should be noted that while four of the 30 tasks in SAS Viya were determined not to be suitable for the Business Analyst persona, no task was considered unachievable due to complex technical skill requirements. Instead, the four tasks – Model Competition, Pipeline Competition, Model Tuning, and Model Comparison – were determined to require additional background knowledge of machine learning and statistics in order to adequately understand model metrics and successfully complete the tasks.

This lack of in-depth machine learning knowledge was found to affect completion in both competitive platforms. In addition, many tasks were determined unachievable in the two competitive environments due to requiring complex programming requirements and data science-specific programming libraries beyond the knowledge of a typical Business Analyst. For some tasks, it was found that SQL-based visual dashboard capabilities could be utilized in the commercial platform, enabling slightly greater utilization of the Business Analyst throughout the life cycle, compared to the non-commercial environment.

Notably, along with a low-code environment and several automated actions, SAS Viya was found to additionally assist the Business Analyst with auto-generated, natural language descriptions of various visualizations and metrics that were not available in the competitive platforms. While in some cases, it was determined that a greater level of knowledge was still required to successfully complete the task, these natural language descriptions lower the overall learning curve for users without in-depth expertise around data and AI.

Improving Productivity with SAS Viya

This evaluation demonstrated that SAS Viya is capable of enabling significant productivity gains compared to competitive data and AI platforms. SAS Viya was found to enable navigation of the data and AI life cycle in less time and with less experienced resources than either competitive platform. Additionally, SAS Viya was found to provide an easy to use, yet powerful tool to complete data and AI tasks. Throughout this evaluation, SAS Viya was found to provide the following productivity benefits:

- Reduced time required to navigate the data and AI life cycle
- Maximized utilization of all users
- Enhanced collaboration between all users
- Flexible support of data sources, code, and no-code approaches
- Minimized time spent on infrastructure and software management

The Futurum Group believes that the findings in this study demonstrate a productivity advantage that organizations can leverage to achieve tangible business benefits, including those outlined below.

Faster Business Decisions

A key finding from this study was the reduction of time spent navigating the data and AI life cycle, achieving the transformation of data to decision with 4.6x greater productivity than competitors on average. Along with increasing productivity of the full life cycle, it was additionally found that SAS Viya provided a productivity boost at each individual stage, with 16x less time spent on Data Engineering roles, 3.5x less time spent on Data Science roles, and 4.5x less time spent completing MLOps roles.

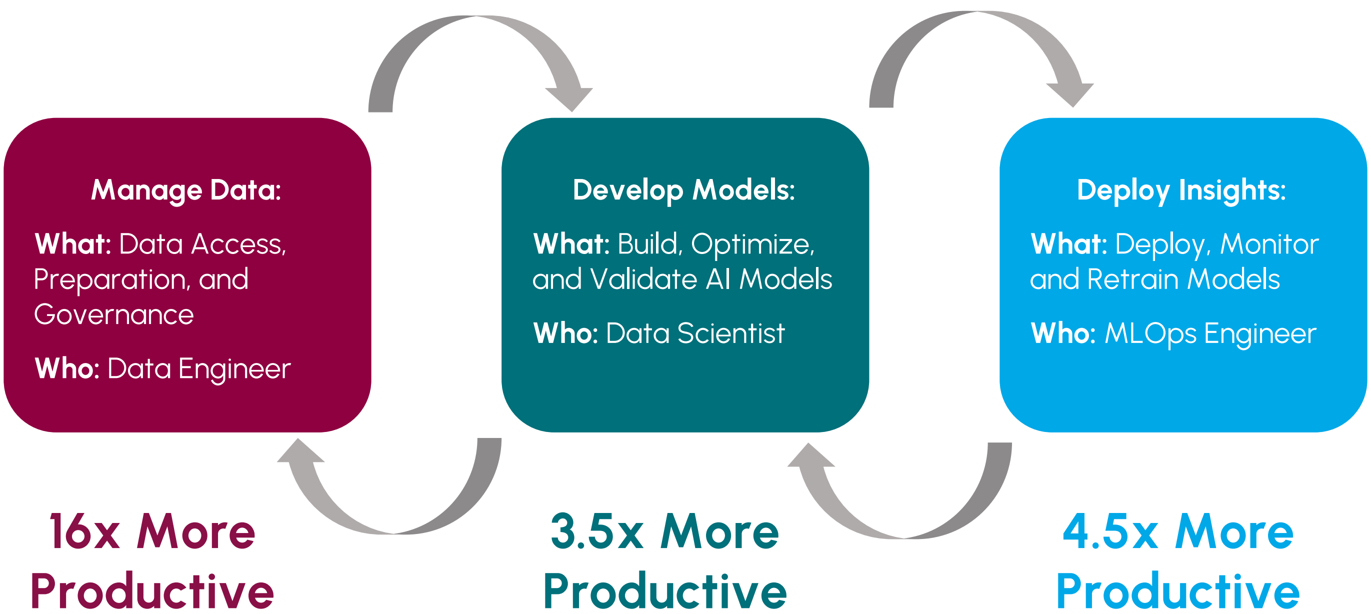


Figure 8: Productivity Throughout Data and AI Life Cycle

It is also crucial to note that these measurements do not represent a one-time increase in productivity. While this study evaluated a single pass through the data and AI life cycle, for most organizations this is an iterative process, requiring ongoing effort at each individual stage. By leveraging SAS Viya, the productivity gains found in this study will compound over time as the life cycle is repeated.

The time savings found in SAS Viya were enabled by simplifying and automating otherwise complex tasks throughout the data and AI life cycle, such as providing automated data scanning, drag-and-drop ETL flows, pre-built model templates, and one-click model deployment. It is also notable that previous Futurum Group performance testing found SAS Viya to provide a significant performance gain over competitors, with a 30x performance advantage, on average. This combination of features and performance not only provides time savings throughout the data and AI life cycle, but ultimately allows organizations to reach critical business decisions faster.

Enhanced Innovation

By achieving greater productivity with SAS Viya, organizations can ultimately achieve greater overall innovation. SAS Viya provides several features that both increase productivity and can be leveraged to achieve greater insights and innovation.

Notably, SAS Viya provides a centralized platform that can connect to a wide variety of data sources, allowing organizations to leverage all of their data as needed. Combining a multitude of data sources with the powerful data analytics and modeling capabilities achievable in SAS Viya presents an opportunity to quickly uncover new insights and innovations.

A key area in which organizations are seeking this type of data-driven innovation is through utilization of new generative AI technologies. While the customer churn predictive analysis completed in this evaluation was not reliant on such technology, The Futurum Group did note the ability of SAS Viya to integrate with and orchestrate GenAI tools. SAS Viya was found to provide a flexible yet practical approach to GenAI, in which new tools could be integrated into the platform to add intelligent decisioning to existing business workflows.

By increasing productivity throughout the data and AI life cycle, SAS Viya can unlock time for data and AI teams to develop new projects and achieve this enhanced innovation. This can be achieved not only through the time savings found in navigating the data and AI life cycle, but also by enabling greater utilization of non-technical resources. As was found through the Business Analyst evaluation, SAS Viya enables a non-technical resource to achieve 86% of tasks throughout the data and AI life cycle. By leveraging such resources, organizations can free up time for expert data and AI practitioners to tackle more innovative tasks where their expertise is invaluable, such as development of new GenAI-enabled business applications.

Cost Savings

The productivity gains achievable with SAS Viya may additionally lead to an economic benefit for organizations. By lowering complexity throughout the data and AI life cycle, SAS Viya was found to enable greater use of non-technical resources. This allows organizations to maximize the productivity of their existing workforce, including non-technical or junior resources, and reduce the operational costs associated with large teams of data and AI experts. For the most complex tasks, SAS Viya provides a more streamlined path to upskilling existing workers than competitive environments.

Additionally, SAS Viya enables data and AI teams to focus on their core tasks, rather than spend costly time and effort managing infrastructure and software dependencies, such as was found in competitive platforms.

Increased Revenue

Ultimately, increasing productivity throughout the data and AI life cycle can enable organizations to operate faster and increase their overall revenue. By achieving data insights faster with SAS Viya, organizations can solve more business challenges and create more data-driven products. In addition, the simple and efficient modelops process found in SAS Viya enables organizations to maintain their models with greater agility, ensuring models maintain value throughout their life.

Competitive Advantage

The combination of these benefits ultimately enables organizations to achieve a significant advantage over competitors who are unable to realize the same level of productivity throughout the data and AI life cycle. The productivity gains found with SAS Viya allow organizations to iterate through the life cycle faster to achieve more insights and develop more useful models. In addition, SAS Viya presents several key features that enable organizations to achieve a competitive advantage, including capabilities that enhance the overall collaboration between data and AI teams and provide built-in data governance capabilities.

Example Scenarios

The results from this evaluation show significant time savings achieved at each stage of the data and AI life cycle; however, the specific impact on productivity and time savings will vary from organization to organization. This evaluation aimed to capture time measurements of a single iteration through common tasks in the data and AI life cycle, which could be further extrapolated to estimate productivity gains in more complex environments.

The specific breakdown of time spent at the Manage Data, Develop Models, or Deploy Insights phases of the data and AI life cycle will be highly dependent on specific factors, including the data utilized, the business challenges addressed, and the data practitioners involved. To provide a greater understanding of how results of this study might apply to various organizations, the productivity gains have been applied to distinct scenarios representing various breakdowns of time spent at each phase.

All scenarios assume 2000 total working hours spent over the course of a year, representative of a single full-time data practitioner working 40 hours a week for 50 weeks, with various breakdowns of time spent. Calculations of time spent in SAS Viya are created utilizing the time reductions found at each phase, as compared to the closest competitive platform. Consideration of additional tasks that may be completed were included and classified as Other. No productivity advantage was applied to these tasks, as they were considered outside the scope of this study.

Scenario A

Scenario A represents the "80/20 Rule", a commonly cited concept stating that data practitioners often find themselves spending 80% of their time managing and cleaning data, and only 20% of time creating models and insights. To approximate this scenario, 80% of time was attributed to Manage Data, while the remaining 20% was distributed evenly between Develop Models and Deploy Insights. For this scenario, it was assumed no time was spent on additional tasks. This scenario demonstrates 89% total time savings.

Task	% of Time	Yearly Time Spent – Competitive (Hrs.)	Yearly Time Spent – SAS Viya (Hrs.)
Manage Data	80.00%	1600	100
Develop Models	10.00%	200	66.67
Deploy Insights	10.00%	200	44.44
Other	0.00%	0	0
Total	100.00%	2000	211.11
		Time Savings	**89%**

Figure 9: Scenario A

The high percentage of time savings found in this scenario is attributed to a large majority of time allocated to the Manage Data phase, in which SAS Viya was found to have its highest productivity gains. While the "80/20 Rule" may not be applicable to all organizations, it is commonly accepted that data cleaning and management tasks often consume a significant amount of time for most organizations.

Scenario B

The second scenario attempts to model a balanced, yet realistic approach based on market research observed by The Futurum Group. This scenario acknowledges that while the Manage Data stage often consumes the largest portion of time in the data and AI life cycle, more time may be spent on additional model development and deployment steps than is allotted in the "80/20 Rule". This scenario also acknowledges that a small percentage of time may be spent on additional tasks, outside of what has been defined as the three core phases of the data and AI life cycle.

In total, this scenario resulted in a 79% time savings.

Task	% of Time	Yearly Time Spent – Competitive (Hrs.)	Yearly Time Spent – SAS Viya (Hrs.)
Manage Data	44.00%	880	55
Develop Models	31.00%	620	206.67
Deploy Insights	22.00%	440	97.78
Other	3.00%	60	60
Total	100.00%	2000	419.45
		Time Savings	**79%**

Figure 10: Scenario B

Final Thoughts

As organizations continue to turn to their data to solve business challenges, the data and AI life cycle will become an increasingly crucial process to drive business success. Despite its critical nature, this process is often hampered by complex, time-consuming tasks and limited resources available to complete them. To overcome these challenges, organizations can turn to new and innovative tools capable of increasing productivity throughout the data and AI life cycle.

This evaluation compared three such approaches, including SAS Viya, a competitive commercial platform, and a non-commercial approach, and found that SAS Viya was capable of dramatically increasing productivity of data and AI practitioners. SAS Viya was found to enable more productive completion of each individual phase of the overall life cycle, as well as lower the barrier for non-technical resources to contribute.

The impact on productivity provided by SAS Viya enables organizations to overcome common challenges faced by data and AI teams, such as limited time and resources, as well as assisting with general business challenges including faster decisions, reducing costs, driving innovation, and increasing revenue. While the total impact on productivity may vary between organizations and their specific business challenges, the findings of this testing are compelling. In addition to providing a significant time advantage when navigating the data and AI life cycle, SAS Viya was found to be easy to use while providing a strong set of built-in features and automations to effectively navigate the data and AI life cycle.

Appendix

Full testing results of all tasks performed in each environment are provided below.

Data Engineering Tasks

	SAS Viya	Commercial	Non-Commercial	Advantage
Data Upload				
Time	1 min	1 min	1 min	None
Complexity	Low	Low	Low	
Coding Required	No	No	Yes	
Achievable by Business Analyst	Yes	Yes	Yes	
Data Profiling				
Time	5.5 min	10 min	10 min	SAS
Complexity	Low	Medium	Medium	
Coding Required	No	No	Yes	
Achievable by Business Analyst	Yes	Yes	Yes	
Data Sensitivity				
Time	1 min	35 min	35 min	SAS
Complexity	Low	Medium	Medium	
Coding Required	No	No	Yes	
Achievable by Business Analyst	Yes	No	No	
Data Quality				
Time	5 min	20 min	20 min	SAS
Complexity	Low	Medium	Medium	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	No	No	
ETL				
Time	6 min	4 Hours	4 Hours	SAS
Complexity	Low	High	High	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	No	No	

Figure 11: Data Engineer Test Results

Data Scientist Tasks

	SAS Viya	Commercial	Non-Commercial	Advantage
Visual Exploration and Insights Discovery				
Time	3 min	207 min	207 min	SAS
Complexity	Low	High	High	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	Yes	Yes	
Visual Exploration - Augmented Analytics				
Time	9 min	29 min	29 min	SAS
Complexity	Low	Medium	Medium	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	Yes	Yes	
Outlier Detection				
Time	10 min	27 min	27 min	SAS
Complexity	Low	Low	Low	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	Yes	Yes	
Model Tuning				
Time	10 min	29 min	29 min	SAS
Complexity	Low	Medium	Medium	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	No	No	No	
Explainability				
Time	46 min	61 min	61 min	SAS
Complexity	Low	Medium	Medium	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	No	No	
Bias Detection				
Time	8 min	31 min	31 min	SAS
Complexity	Low	High	High	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	No	No	

	SAS Viya	Commercial	Non-Commercial	Advantage
Model Reports				
Time	3 min	6 min	1 Hour	SAS/Commercial
Complexity	Low	Low	Medium	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	Yes	No	
AutoML				
Time	7 min	7 min	20 min	SAS/Commercial
Complexity	Low	Low	Medium	
Coding Required	No	No	Yes	
Achievable by Business Analyst	Yes	Yes	No	
Pipeline Competition				
Time	12 min	32 min	32 min	SAS
Complexity	Medium	High	High	
Coding Required	No	No	No	
Achievable by Business Analyst	No	No	No	
Model Registration				
Time	2 min	21 min	30 min	SAS
Complexity	Low	Low	Medium	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	Yes	Yes	
Project Insights Report				
Time	2 min	2 min	30 min	SAS/Commercial
Complexity	Low	Low	Medium	
Coding Required	No	No	Yes	
Achievable by Business Analyst	Yes	Yes	No	
Sharing Projects				
Time	1 min	10 min	10 min	None
Complexity	Low	Low	Low	
Coding Required	No	No	No	
Achievable by Business Analyst	Yes	Yes	Yes	

Figure 12: Data Scientist Test Results

MLOps Engineer Tasks

	SAS Viya	Commercial	Non-Commercial	Advantage
View Models in Central Repository				
Time	1 min	1 min	3 min	None
Complexity	Low	Low	Low	
Coding Required	No	No	No	
Achievable by Business Analyst	Yes	Yes	Yes	
Generation of Model Files				
Time	30 sec	30 sec	5 min	None
Complexity	Low	Low	Medium	
Coding Required	No	No	No	
Achievable by Business Analyst	Yes	Yes	Yes	
Model Comparison				
Time	5 min	5 min	5 min	None
Complexity	Low	Low	Low	
Coding Required	No	No	No	
Achievable by Business Analyst	No	No	No	
Model Scoring				
Time	5 min	15 min	20 min	SAS
Complexity	Low	Medium	Medium	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	No	No	
Model Deployment				
Time	1 min	1 min	3 min	SAS/Commercial
Complexity	Low	Low	Low	
Coding Required	No	No	Yes	
Achievable by Business Analyst	Yes	Yes	Yes	
Monitor Model performance				
Time	15 min	2 hours	2 hours	SAS
Complexity	Low	High	High	
Coding Required	No	No	Yes	
Achievable by Business Analyst	Yes	Yes	Yes	

	SAS Viya	Commercial	Non-Commercial	Advantage
Model Retraining				
Time	5 min	35 min	35 min	SAS
Complexity	Low	Medium	Medium	
Coding Required	No	Yes	Yes	
Achievable by Business Analyst	Yes	No	No	
Versioning				
Time	1 min	1 min	1 min	SAS
Complexity	Low	Low	Low	
Coding Required	No	No	Yes	
Achievable by Business Analyst	Yes	Yes	Yes	
Alerting				
Time	1 min	1 min	3 min	SAS/Commercial
Complexity	Low	Low	Low	
Coding Required	No	No	No	
Achievable by Business Analyst	Yes	Yes	Yes	
Scheduling				
Time	5 min	5 min	5 min	None
Complexity	Low	Low	Low	
Coding Required	No	No	No	
Achievable by Business Analyst	Yes	Yes	Yes	

Figure 13: MLOps Engineer Test Results

Contributors

Data Engineer

Russ Fellows – VP of Labs, The Futurum Group

A veteran technologist, with a combination of technical, marketing, and analytic skills, from over 30 years working in engineering, IT management, marketing strategy, and as an industry analyst. An experienced leader, manager, and subject matter expert in multiple domains, Russ works with practitioners across technologies and industries.

Data Scientist

Frederic Van Haren – Strategic Advisor - Data Science and Analytics, The Futurum Group

Frederic Van Haren is the CTO at HighFens Inc. and serves as a Strategic Advisor for The Futurum Group's Data Science and Analytics testing. Frederic is known for his insights into the AI, Big Data, HPC, and storage industries. He has over 20 years of high-tech experience and has built several large HPC and AI environments from the ground up.

MLOps Engineer

Brian Martin – AI and Data Center Lead, The Futurum Group

After three decades leading innovation in distributed system design and performance across all aspects of IT in companies large and small, Brian leads AI and Datacenter Performance for The Futurum Group.

Business Analyst

Mitch Lewis – Research Analyst, The Futurum Group

With an extensive background in computer science and data science, Mitch brings deep technical knowledge of data storage, data management, and AI technologies. Before joining The Futurum Group, Mitch served as an industry expert in information management with Evaluator Group and previously led cloud implementations during his time at Oracle.

Important Information About this Report

CONTRIBUTORS

Mitch Lewis
Research Analyst | The Futurum Group

Russ Fellows
VP Labs | Signal65

Brian Martin
AI and Data Center Lead | Signal65

Frederic Van Haren
Strategic Advisor - Data Science and Analytics | The Futurum Group

PUBLISHER

Daniel Newman
CEO | The Futurum Group

INQUIRIES

Contact us if you would like to discuss this report and The Futurum Group will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "The Futurum Group." Non-press and non-analysts must receive prior written permission by The Futurum Group for any citations.

LICENSING

This document, including any supporting materials, is owned by The Futurum Group. This publication may not be reproduced, distributed, or shared in any form without the prior written permission of The Futurum Group.

DISCLOSURES

The Futurum Group provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

ABOUT SAS

SAS provides trusted data and AI solutions to global customers, delivering the knowledge they need in critical moments. The company drives bold innovation across industries. Visit their website to learn more about how SAS can support your business.

ABOUT THE FUTURUM GROUP

The Futurum Group is an independent research, analysis, and advisory firm, focused on digital innovation and market-disrupting technologies and trends. Every day our analysts, researchers, and advisors help business leaders from around the world anticipate tectonic shifts in their industries and leverage disruptive innovation to either gain or maintain a competitive advantage in their markets.

CONTACT INFORMATION

The Futurum Group LLC I futurumgroup.com I (833) 722-5337 I

114060_G281740.0924