Paper 4917-2020

Enhancing Academic Teaching of Data Science with SAS® Viya® for Learners

Stefan Dimitrov Stoyanov, University of Surrey

ABSTRACT

SAS® Viya® for Learners offers many advanced tools for studying data science. I found the fast cloud suite especially helpful to overcome obstacles in academic teaching that hinder novice learners. As a newcomer to data science lacking strong statistics knowledge, it was difficult to promptly read large textbooks or absorb in-depth knowledge from slides while in school, with limited hours for consultations with lecturers and complicated software. My paper discusses the features of SAS Viya for Learners that helped me to overcome these obstacles: interactive courses in statistics, programming, and advanced analytics, videos, summaries, examples, practices, case studies, papers, documentation, online chat with educational experts, webinars, and communities. The advantages of SAS Viya for Learners over IBM SPSS® Modeler and R made it much easier to comprehend machine learning and appreciate the entire analytics life cycle thanks to Model Studio and other solutions. I found **it enabled students' collaboration, combining wel**l-structured visual analytics workflow with SAS® and open-source code, parameters that are easy to set up, automated hyperparameter tuning, assessment visualizations, and best model selection. By including SAS Viya for Learners in academic curriculums, I strongly feel that educators could teach more creatively to speed-**up students' deep data science comprehension, providing better**-prepared professionals to bridge the gap in data science skills. Join me for my presentation to find out more.

INTRODUCTION

Data insights become crucial for the development of organizations from almost every business sector and research field. To respond to the growing demand for professionals with data science skills, many universities are doing their part to ensure that students with deficiencies in mathematics, statistics, and programming are taught essential data analytics skills. However, still many university curriculums are designed primarily for students with a substantial STEM background. Existing academic programs must be updated to be more responsive to the learners' needs. The American Statistical Association (ASA) recommends modernized curriculums to teach statistical thinking for problem-solving and decision making; focus on conceptual understanding; integrate real data with a context and purpose; foster active learning and use appropriate technology to explore concepts and analyze data (Committee, 2016). This paper demonstrates how SAS Viya for Learners suite addresses these requirements by offering a variety of advanced tools and useful resources to complement and advance academic teaching for novice learners in data science significantly.

The remainder of the paper is organized as follows. The section "The Importance of Data Science Academic Programs" evidences the need and the benefits from high-quality academic curriculums having the potential for bridging the increasing data science skills gap. The next section, "Common Obstacles in Academic Teaching for Novice Data Science Students", examines characteristic limitations of academic curriculums that hinder

newcomers to data science from absorbing the taught concepts more promptly and efficiently. The section "Why SAS® Viya® for Learners Improves Data Science Academic Programs" demonstrates the advantages of the powerful, unified and open educational platform for teaching advanced analytics concepts over other leading software options such as IBM SPSS Modeler and R. Further, the chapter highlights the rich pool of learning and teaching resources of SAS Viya for Learners which are advantageous for teaching data science to novice students. The last section of the paper "How SAS Viya for Learners Could be Even Better with Data-Driven Apps" accentuates how SAS Viya for Learners can reach the next level of data science education by empowering students with skills to complete the last mile of analytics quickly, efficiently, and at scale by building highly tailored data apps in a democratized way.

THE IMPORTANCE OF DATA SCIENCE ACADEMIC PROGRAMS

Comprehensive academic curriculums of data science disciplines are essential for delivering high-quality education to respond to the emergent needs of corporations and academia for professionals able to extract value from vast amounts of information.

BOOMING DEMAND FOR DATA SCIENCE SKILLS

The demand for specialist data skills more than tripled over five years (+231%) in the UK alone (The Royal Society, 2019). Data Science and Analytics (DSA) job listings in the USA are expected to increase from 364,000 openings in 2016 to approximately 2,720,000 in 2020 (Columbus, 2017). The figure below shows that the data scientist role is with the highest projected growth and among the hardest to be fulfilled:
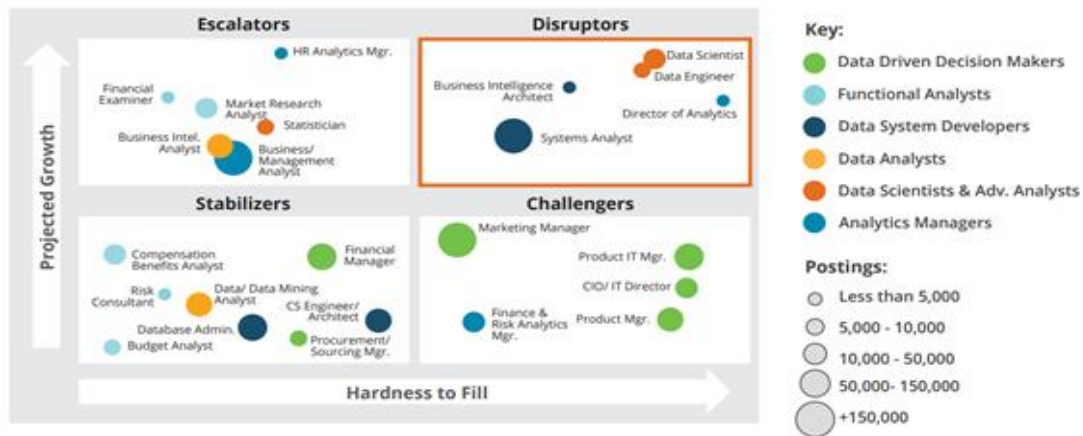

Figure 1. DSA Jobs Matrix DSA Jobs Matrix (Burning Glass Technologies, 2017)

BRIDGING THE DATA SCIENCE SKILLS GAP WITH HIGH-QUALITY ACADEMIC PROGRAMS

Despite the growing demand, over 50% of the UK tech industry employers explain that candidates do not possess the right technical skills (Robert Walters Group, 2018). Therefore, as you can see from the figure below, it is not a surprise that 98% of Data Scientists and Advanced Analyst positions require higher education degrees (Columbus, 2017).

Figure 2. Level of Education Required by DSA Job Postings in the USA (PwC, 2017)

Business relies on comprehensive academic curriculums to help them bridge the gap in data science skills. Effective teaching programs in data science have to provide multiple holistic benefits:

- Develop students as highly employable analytics experts needed by corporations to deliver change management and transition to data-driven businesses in industry 4.0.

- Prepare students for modern scientific research.

- Equip students with up-to-date practical skills for performing real-world applications.

- Boost students' creativity, prepare, and stimulate them to establish start-up businesses.

To meet these expectations, high-quality educational curriculums must be responsive to workforce needs and equip students with a broad range of specialized analytical skills and tools, such as Data Mining and Machine Learning, SAS, SQL, R, Python (Burning Glass Technologies, 2017). Also, **students have to be able to acquire "newer, emerging skills" like** business intelligence (BI) & Data Management, considered by 41% of UK Chief Technology Officers (CTOs) to be among the top required skills (Stoyanov, 2019).

Further, an excellent academic teaching program must equip students with highly valued by the industry non-technical transferable skills in data science. Such skills include communication and storytelling, analytical problem-solving, strategic, proactive, creative, innovative, and collaborative skills (Burning Glass Technologies, 2017). Besides, a well-constructed academic curriculum must provide students with substantial domain knowledge. Moreover, academic teaching must be motivational to create data scientists who are passionate about the business domain in which they will work.

UNPARALLELED BENEFITS OF FULL-TIME ACADEMIC PROGRAMS FOR DATA SCIENCE STUDENTS

In addition to these characteristics of effective data science curriculums, I have discovered that full-time studying programs at university can provide you with many benefits to enhance your development:

- Interaction with coursemates from various backgrounds enriches your perspectives on the taught topics and broadens your horizons.

- Working on individual and group coursework helps you to build technical and soft skills.

- You also have the opportunity for engagement and discussions during lectures and labs.

- Furthermore, you have the opportunity for a one-to-one consultation with your lecturers. You can receive valuable practical advice from recognized academics and successful practitioners to foster your academic studies and career prospects.

- Academic teaching builds your ability to conduct independent research. You also develop skills for writing and presenting academic and business reports.

- Moreover, at University, you can interact with students and researchers from many other academic departments and various levels of expertise. This diversity provides endless opportunities for collaboration in highly specialized and interdisciplinary teams. You can benefit from the advantages of sharing knowledge, curiosity, and experience, which, in turn, spark innovation.

- Last but not least, the possibility for a placement enables you to further comprehend and put in practice the taught data science concepts to solve real business problems.

COMMON OBSTACLES IN ACADEMIC TEACHING FOR NOVICE DATA SCIENCE STUDENTS

Although academic programs can provide a number of unparalleled benefits, many of them still have characteristic limitations hindering novice students to more quickly and thoroughly comprehend the wide variety of data science topics during the short academic semesters.

Within my MSc Business Analytics program, I encountered and observed a number of different learning obstacles. Many of these hurdles are not unique to data science programs such as: long reading lists, presentations lacking in-depth technical information, and having limited contact and office hours with educators.

However, in addition to the above obstacles, there are a number of others that I experienced and which I believe are particularly prominent for data science programs:

- Extensive and challenging to learn textbooks.

  Many books impart data science knowledge with the assumption that you already have a good understanding of applied statistical and mathematical concepts. Such prerequisites hinder you as a novice learner to obtain a sound comprehension of the taught concepts.

  For example, this prior knowledge deficiency made it difficult for me to comprehend quickly and sufficiently, the advanced concepts taught by the Econometrics module (University of Surrey, 2018) during the first semester of my MSc Business Analytics course.

- Lack of alternative concise learning sources.

  As a novice data science learner struggling to obtain quickly sufficient knowledge from large books or slide presentations, I started searching for compact and interactive online learning material with good quality. However, it used to take me a lot of time until I came across a video tutorial or other interactive tool to satisfy my learning needs.

- Outdated or complicated teaching software.

  R, as an extensible programming language, enables a knowledgeable and advanced data scientist to construct their methods for data analysis. However, the lack of any previous coding experience and knowledge about programming logic made it overwhelming for me to implement in R the concepts taught within the Machine Learning Module (University of Surrey, 2018). As a novice learner, I had to spend more time learning to work with R, rather than to comprehend the data science knowledge. Researchers like John Harraway also point out that using R to introduce data science concepts to newcomers is too complicated and frustrating for them (Harraway, 2012).

  Tools for visual data mining and machine learning provide beginners with more frictionless initial comprehension and hands-on experience to put in practice the data science taught theory. However, they must provide a modern and effective user experience and guidance through the analytics lifecycle.

- Increased levels of stress caused by the factors mentioned above.

As a result of all these hurdles, I felt increased levels of stress and anxiety. High academic stress among university students is recognized as a significant problem nowadays by many educational researchers too. Such a problem very often lowers **student's performance and prevents them from discovering and realizing their full** potential. Researchers point out that inadequate resources, the vastness of syllabuses, and expectations for long study hours are substantial reasons for increased stress levels experienced by students (Reddy, et al., 2018).

Educational programs need to evolve to overcome all these characteristic obstacles in data science teaching. Modern academic curriculums must introduce more flexible and innovative methods and tools to enable novice learners to absorb the taught concepts more promptly and efficiently.

WHY SAS® VIYA® FOR LEARNERS IMPROVES DATA SCIENCE ACADEMIC PROGRAMS

When I started learning SAS Viya, I was surprised how easy it was for me to absorb the knowledge taught in the SAS Training Programs. They changed my learning experience most positively and boosted my confidence.

I realized that SAS Viya for Learners provides you with the tools to assist you in overcoming the typical limitations of academic programs teaching novice data science students. I firmly believe that by following the well-structured teaching approach provided by SAS Viya for Learners, anyone motivated and dedicated can become a successful data scientist. Here is why:

FREE AND POWERFUL SUITE OF SAS® ANALYTICS SOFTWARE

SAS Viya for Learners is a full free suite of cloud-based software that supports the entire analytics life cycle – from data to discovery to deployment (SAS Institute, 2020). It lets you access SAS artificial intelligence (AI) and machine learning (ML) analytics tools (Khan, 2019).

SAS Viya for Learners provides:

- Powerful analytics.

  SAS Viya is an in-memory analytics platform that supports faster elastic, scalable, and fault-tolerant processing for huge amounts of data to address complex analytical challenges.

  Faster analytics provides you as a student with more time to study more real business use cases. In this way, you, as an educator, have more opportunities for teaching analytics and critical thinking skills.

- Unified and open analytics environment.

  A unified analytics environment with a standardized code base supports programming in SAS as well as open-source like Python and R.

  This integration provides you as a student with endless analysis opportunities, delivering a truly open experience. Simultaneously, SAS Viya empowers you as an educator to be more creative by allowing you to teach a rich pool of SAS and open source tools.

ADVANTAGES OVER THE MAIN ALTERNATIVE SOFTWARE TOOLS

Most academic modules that teach data mining and machine learning include coursework that you, as a student, have to accomplish. You need to use specialized software tools for performing data science tasks.

During the first and second academic semesters of my MSc Business Analytics course, I learned to use software for advanced analytics like IBM SPSS Modeler, R, and RStudio®.

IBM SPSS Modeler is a visual data science and machine learning solution (IBM, 2020). Like SAS Viya, it is used to build predictive models by applying statistical and data mining algorithms without the need for programming.

R is an open-source programming language and environment, which provides integrated software facilities for data calculation, graphical display, as well as a large collection of intermediate tools for data analysis (The R Foundation, 2020). R is highly extensible via packages. The R distribution includes eight packages. Many more are available through the Comprehensive R Archive Network (CRAN), covering a wide range of modern statistics (The R Foundation, 2020).

Learning IBM SPSS Modeler and R enabled me to begin building machine learning models to gain insights from data for effective business decisions. However, they have limitations that hinder you as a novice learner to quickly and thoroughly understand and implement in practice the whole analytics lifecycle. Here are some of them:

Limitations of IBM SPSS® Modeler

- Clumsy user experience design (UX) and outdated graphical user interface (GUI).

  The UX of SPSS Modeler seems to me to provide too many menus and levels of submenus. Also, many of the analytical tasks, for example, for data exploration, data preparation, or model assessment, are split into too many nodes. You have to open them in separate windows to set up the parameters. Therefore, you need more time to construct your analytics workflow. Also, there is a risk for you, as a novice learner, to miss some critical assessments in your project. Also, the graphics are static and with a very old style. These constraints limit user experience, make the journey to solve a business problem clumsier, and provide less intuitive and seamless guidance through the entire analytics lifecycle.

- At university, we had to perform a lot of tasks manually, which takes time. For example:
  - tuning the hyperparameters of a model,
  - building different assessment graphs,
  - comparison of models,
  - determination of the best performing model.

Limitations of R

- Moreover, coding an entirely new machine learning application from scratch can be time-consuming (Stoyanov, 2019).

- **R lacks simple tools for instant team's communication and collaboration.**

  R allows team collaboration via many channels like e-mails, listserves, bulletin boards, package creation and sharing, and version control in github.com, dropbox, or CRAN (Ross, 2013). You can also integrate Git in R studio (Gillespie & Lovelace, 2016).

However, such collaboration approaches require you to have some advanced coding skills. Moreover, these channels do not support an easy and user-friendly way for instant chat and discussion. Their characteristics limit your ability as a novice learner to implement group ML coursework. The opportunity to share knowledge and skills and have a transparent and instant discussion with your peers in a well-structured manner boosts and speeds-up your data science comprehension.

- Technical dept of R open-source language.

  One of the common problems with open-source languages like R is the technical dept that occurs over time (Smith, 2020). My coursemates and I experienced it too while working on our ML coursework at the university. You start building your ML project within one version of R and RStudio. As you progress, you decide to add an R package for implementing a particular analytics task. Unfortunately, it turns out it is not maintained and is not updated to comply with the R version you are using. Moreover, not a long time after you have started developing your coursework, a new R version is released. You share your coursework with someone who uses the latest R version, and your project gives errors in their environment.

- R programming language is known for having lesser speed and consuming more computational power (DataFlair, 2019). The design of the R language can cause problems in manipulating large datasets. The slower speed of R results in more time needed for the accomplishment of an ML coursework.

- In many real cases, you have to deploy models you build in R in other languages.

  Therefore, you need to have even more advanced coding skills in many more programming languages. As a novice data science learner, you need to study tools that will empower you to deploy your ML models in any environment easily without the need for complicated and time-consuming coding. Such tools will enable you to get a sound understanding of the importance and the processes which put an ML model into production in a much earlier stage of your studies.

All these common software limitations of SPSS Modeler and R hinder you as a student to implement in practice the whole analytics lifecycle within coursework.

Therefore, in many cases, academic coursework designed to be performed with only SPSS Modeler or R requires students to perform only analytical tasks related to data preprocessing, building, and assessment of machine learning models. This was the case with the coursework I had to accomplish at university.

However, other crucial phases of the analytics lifecycle, such as model deployment into **business operations, often are neglected or even left out of a coursework's sco**pe. An ML model adds business value only when put into production. The omission of the deployment phase from coursework narrows your prospect as a novice learner to quickly and thoroughly understand all stages of the analytical model life cycle within a taught academic module. Moreover, recent studies show that 80-85% of enterprises have problems with ML deployment (DataFlair, 2019). More than 60% of all ML and AI models are never operationalized. Only 35 % of organizations fully deploy analytics models (Berman, 2019). Of those that are deployed, 90% took more than three months to be put into production, and 40% took seven months (Hinchcliffe, 2019). These figures make up-to-date ML deployments skills even more important for enterprises.

Therefore, I believe that the above-discussed limitations in the academic teaching process hinder your development as a well-rounded, employment-worthy data scientist.

The SAS Viya For Learners Advantage

I started using SAS Viya for Learners during my placement at Boemska. The "Machine Learning Using SAS Viya" course provided me with the opportunity to go through all the steps of the analytical lifecycle predicting a telecom customer churn (SAS Institute Inc., 2019).

Performing coursework with SAS Viya for Learners provides you with a better Machine Learning introduction. You benefit from exhaustive guidance, which enables you, as a novice student, to better, faster, and more smoothly comprehend the taught concepts, and gain solid practical skills to accomplish the entire ML analytics lifecycle.

This paper explores the characteristics and advantages of the SAS® Visual Data Mining and Machine Learning (SVDMML). SVDMML consists of data mining tools that enable you to explore data, build, and put into production predictive models. Among others, within SAS Viya for Learners, you can use SAS® Data Explorer, SAS® Data Studio, SAS® Visual Analytics, SAS® Model Studio, and SAS® Model Manager solutions to accomplish your ML coursework.

You can use SAS Data Explorer to import and load data into memory, manage, and familiarize with data (SAS Institute Inc., 2020). SAS Data Studio enables you to prepare data easily using an intuitive point-and-click visual self-service interface (SAS Institute Inc., 2020). With SAS Visual Analytics, you can quickly identify relationships and trends in your data through smart visualizations, automated and augmented analytics, and interactive dashboards (SAS Institute Inc., 2020). SAS Model Studio provides you with data mining capabilities to build ML predictive models (SAS Institute Inc., 2020). SAS Model Manager helps you to deploy and govern ML models seamlessly (SAS Institute Inc., 2020).

Here are some of the essential advantages of SVDMML:

- Beautiful, user-friendly interface and advanced UX design.

  SAS Viya provides you with modern UX design. It has a beautiful and simple design and navigates you seamlessly through the whole analytics lifecycle (Stoyanov, 2019). As you can see from the display below, all your projects are well-organized in one place in SAS Drive. You create or open your analytics projects from there:
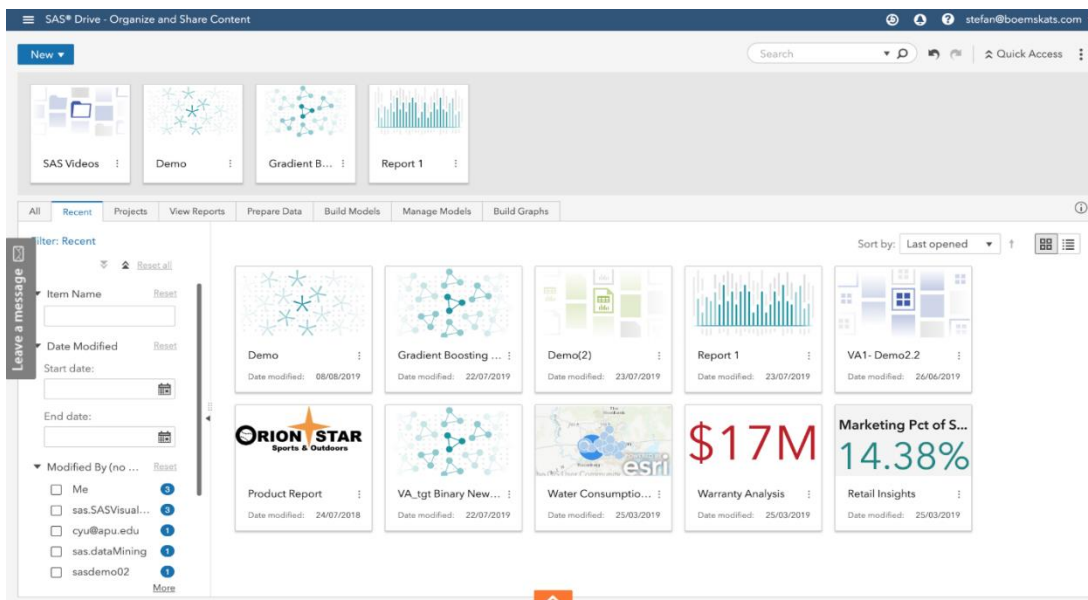
Display 1. SAS® Drive

- A well-structured and guided analytical workflow.

  You can view at a glance the whole well-structured analytical workflow in a single **interactive diagram, called "pipeline". A node in the pipeline represents each analytical** task. You can drag and drop a node from the Nodes pane (on the left part of the display below) into the pipeline to include a new analytical task in your Machine Learning predictive model.
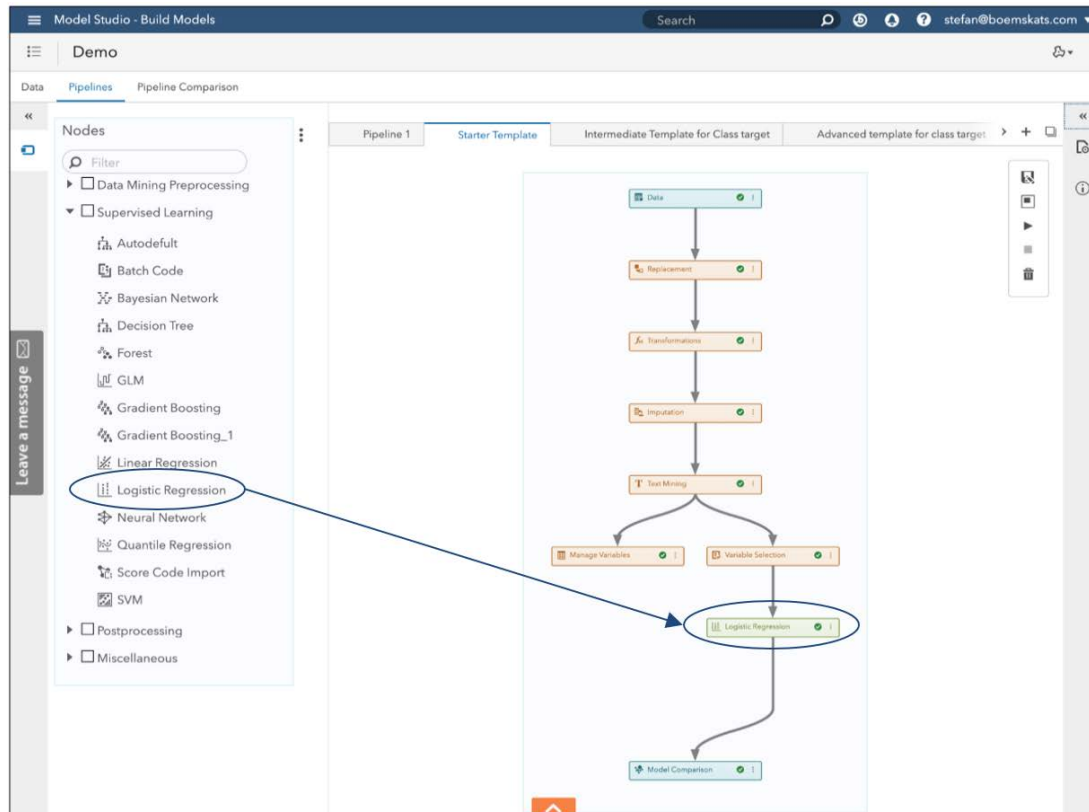
Display 2. Pipeline in SAS® Model Studio

- Easy to set up parameters of the analytical tasks.

  Once you add a node (analytical task) to your workflow, in minutes you can set up its parameters in the properties pane which is at the right side on the same screen where your pipeline is, as shown on the following display:
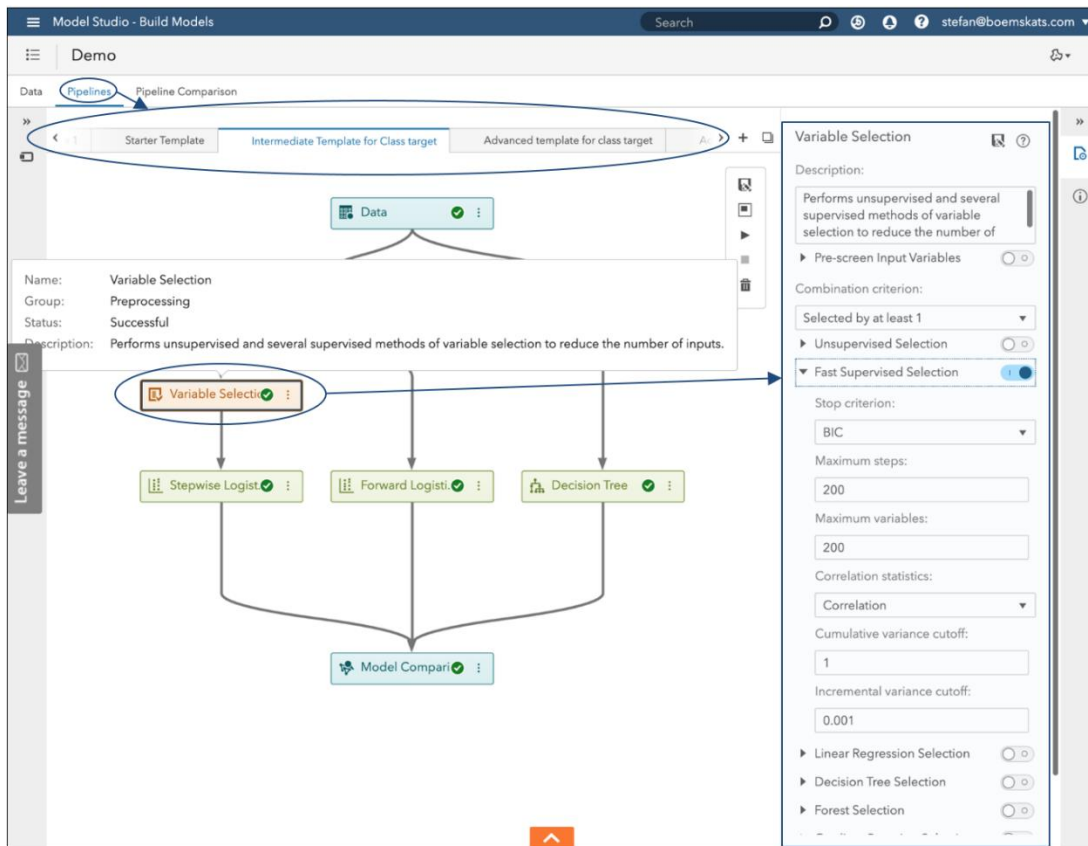
Display 3. Setting Up the Analytics Tasks in Model Studio

That makes the navigation of the whole workflow while building a data mining model easy and quick. You, as a novice learner with limited technical skills, could master the workflow only one day after you start using Model Studio. Moreover, you can build as many pipelines and models as you need in your project and quickly navigate across them.

- Sophisticated integration of advanced analytics techniques.

Model studio enables you to seamlessly integrate different advanced analytics tasks in the pipeline while developing your models. For example, you can easily combine and add both structured data and unstructured data to your model. With the text mining node, you transform unstructured text variables into numeric predictors in minutes. Here you can see the integration of the text mining node in the pipeline and its settings pane:
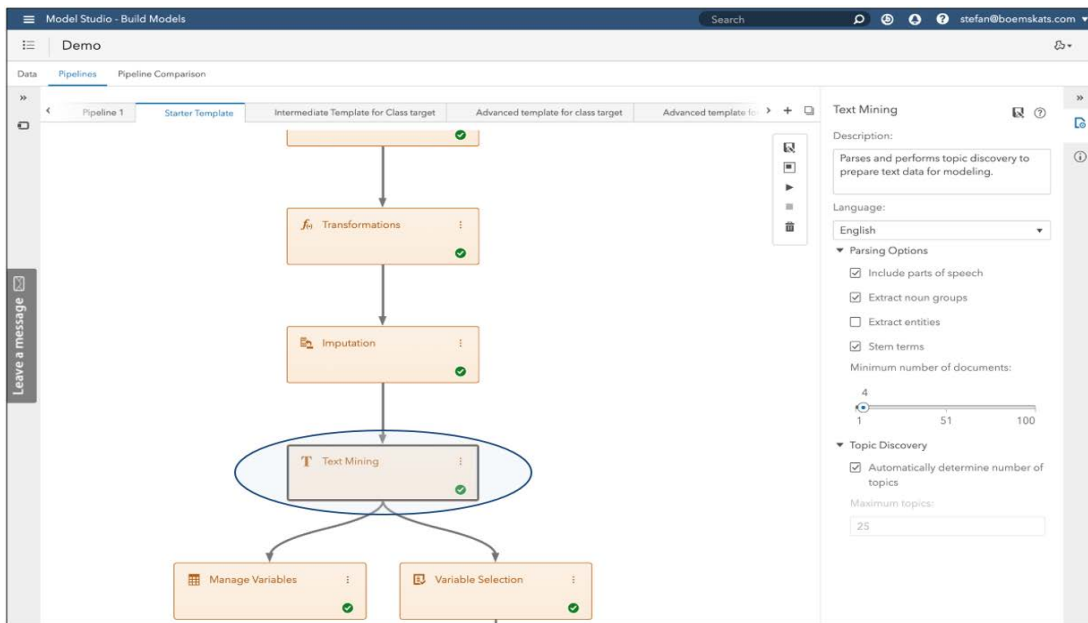
Display 4. Incorporation of the Text Mining Node in Model Studio

- Integration of R models in the analytics pipeline.

  The standardized code base of SAS Viya for Learners supports programming in R. This open-source integration enables students to combine the advantages of SAS with these of R. In this way, SAS Viya for Learners broadens your learning opportunities and equips you with the skills to mix and apply more data science tools. You can explore consistent and governed code. You can easily add an R generated model to your pipeline and automatically compare it with the rest of the models in your ML coursework project. In this way, SAS Viya for Learners empowers you to be more creative when trying to build the best possible performing model and to solve more efficiently a business problem.
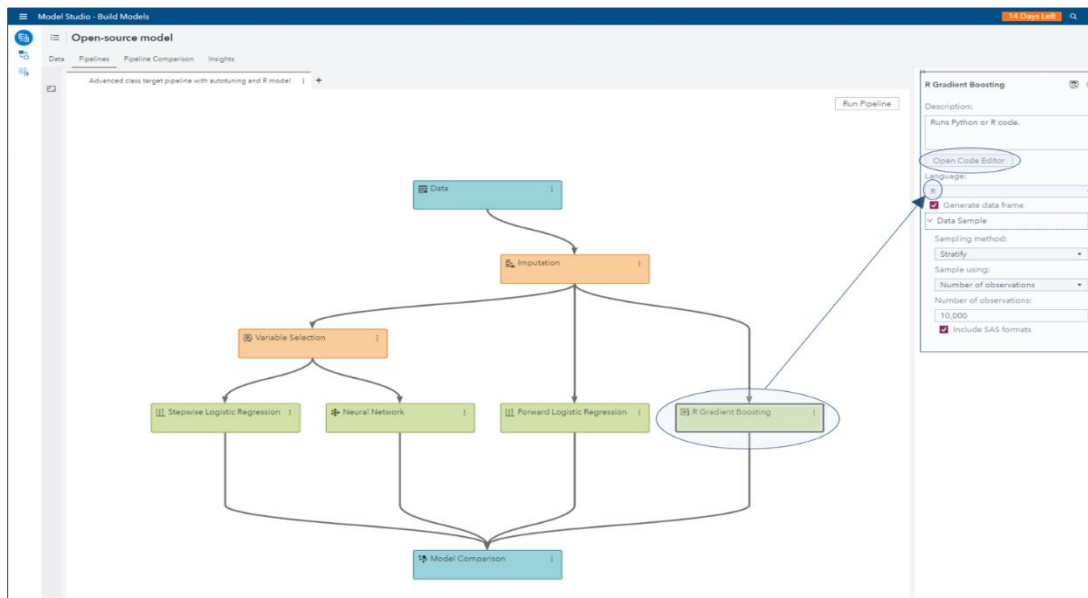
Display 5. Incorporation of the Open-Source Code Node in Model Studio

- Automatic hyperparameter tuning.

  This feature helps you to find the optimal hyperparameters of a Machine Learning algorithm you use to build a model in your pipeline (Stoyanov, 2019). The autotune feature saves you lots of time when you search for the best settings to create the best possible model. Here is an example of the autotuning feature:
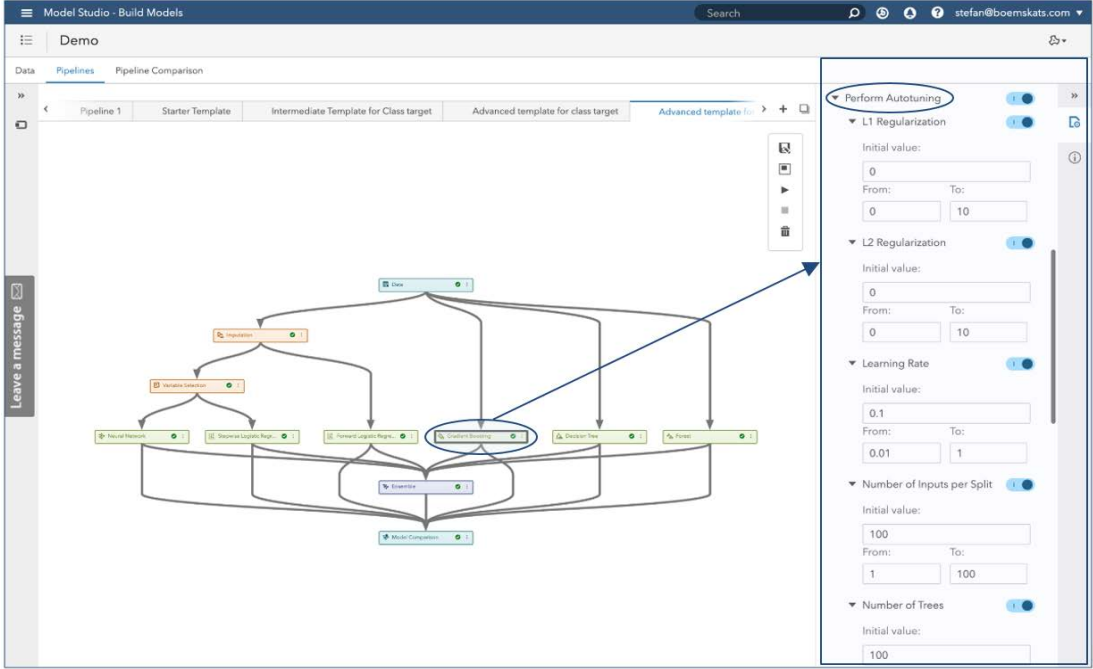
  Display 6. Automatic Hyperparameter Tuning Options for the Gradient Boosting Model in Model Studio

- Automated assessment visualizations of the predictive models.

  Once you have built your model, Model Studio automatically creates many interactive, beautiful, and comprehensive assessment visualizations in graphs and tables. They provide many fit statistics that you can use as a benchmark to assess and compare **models' performance. As you can see below, they are presented in a well**-organized manner:
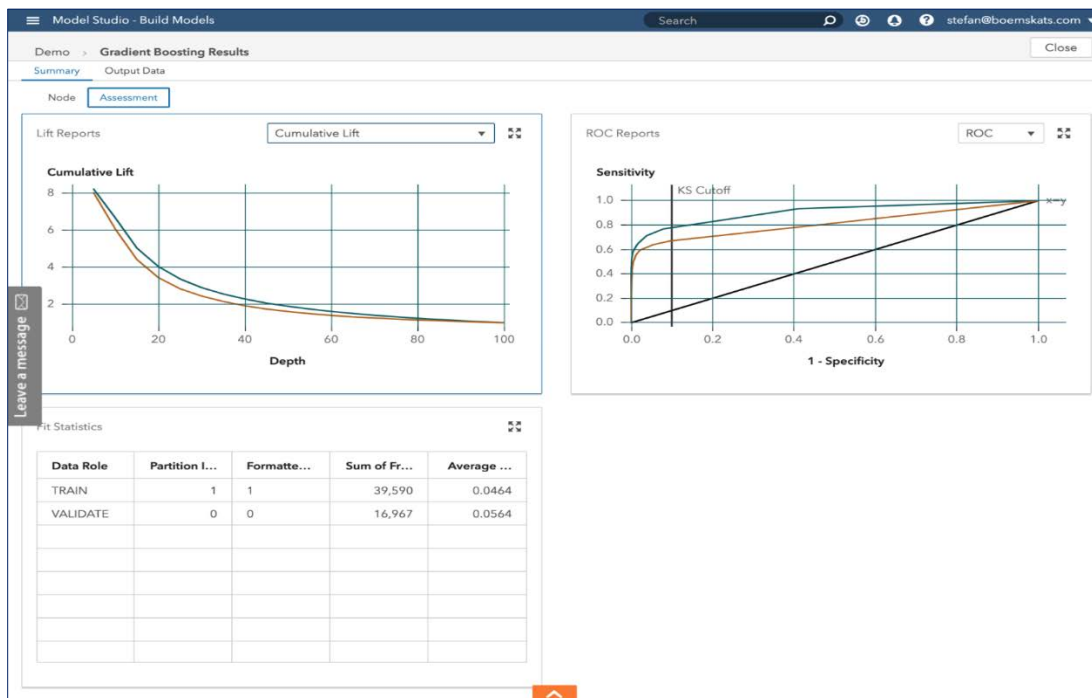
Display 7. Automated Assessment Visualizations in Model Studio

This automation saves you time and helps you to focus on the vital part of your coursework: to quickly, efficiently, and reliably assess the performance of your models.

- Seamless integration of many models and instant choice of the best model in the project.

Model Studio automatically chooses the best model among all models you have built and tested in your project. In a single pipeline, Model studio automatically adds a 'Model Comparison' node as shown below:
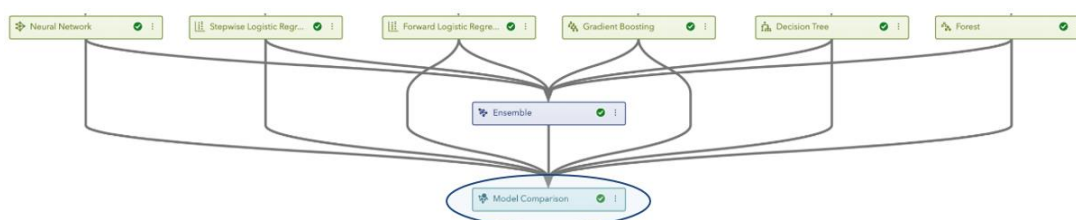
Display 8. Model Comparison Node in Model Studio

It compares the performance of all models in the pipeline, ranks them upon various criteria, and chooses the champion model, as shown below:

Display 9. Model Comparison Fit Statistics Assessment Table in Model Studio

Champion	Name	Algorithm ...	KS (Youden)	Misclassific...	Root Avera...	Average S...	Sum of Fre...	Multi-Class...	Gini Coeffi...
▶ 🔖	Gradient Boosting	Gradient Boosting	0.6049	0.0572	0.2290	0.0524	16,967	0.2106	0.6506
	Forward Logistic Regression	Logistic Regression	0.5909	0.0691	0.2460	0.0605	16,967	0.2366	0.6409
	Ensemble	Ensemble	0.5904	0.0664	0.2443	0.0597	16,967	0.2315	0.6435
	Stepwise Logistic Regression	Logistic Regression	0.5899	0.0693	0.2464	0.0607	16,967	0.2370	0.6369
	Neural Network	Neural Network	0.5882	0.0704	0.2472	0.0611	16,967	0.2378	0.6396
	Forest	Forest	0.5318	0.0951	0.2756	0.0759	16,967	0.2756	0.6096
	Decision Tree	Decision Tree	0.4887	0.0944	0.2924	0.0855	16,967	0.6078	0.5286

Then, Model studio automatically chooses the champion of the champion models across all pipelines in your project:

Display 10. Pipeline Comparison Table in Model Studio

	Champion ▼	Registered	Challenger	Name	Algorithm Name	Pipeline Name	KS (Youden)	Sum of Frequencies
☑	▶ 🔖	☑		Gradient Boosting	Gradient Boosting	Advanced template fo...	0.605	16967
☐				Forward Logistic Regr...	Logistic Regression	Intermediate Templat...	0.591	16967
☐			📌	Gradient Boosting	Gradient Boosting	Tree Based	0.586	16967
☐				Neural Network	Neural Network	Neural Network	0.576	16967
☐				SVM	SVM	Support Vector Machine	0.572	16967
☐				Logistic Regression	Logistic Regression	Starter Template	0.553	16967

- An added layer of transparency and interpretability.

  Machine Learning algorithms like the Support Vector Machine (SVM) are widely **considered as "black boxes" as they are almost** impossible to interpret (Bathaee, 2018). In many cases, they are valued for being able to predict a highly accurate outcome from a model. However, many practitioners cannot use them to explain the relationship between the input variables and the target variable.

  The IML R package can be used to interpret ML models (Molnar, 2019). However, it requires you to have sound coding skills and knowledge of the agnostic methods (Molnar, 2020) to choose the appropriate technique and comprehend the results.

  Model Studio's automated model interpretability feature changes this. It enables even you as a novice learner with less experience and technical skills to quickly and easily find some meaningful insight into the relationships between the predictors and the target **variable in "black boxes" as SVM. It also helps you to quickly and seamlessly** comprehend a variety of interpretability techniques. Here is a view of the Model Studio's automated model interpretability settings pane:
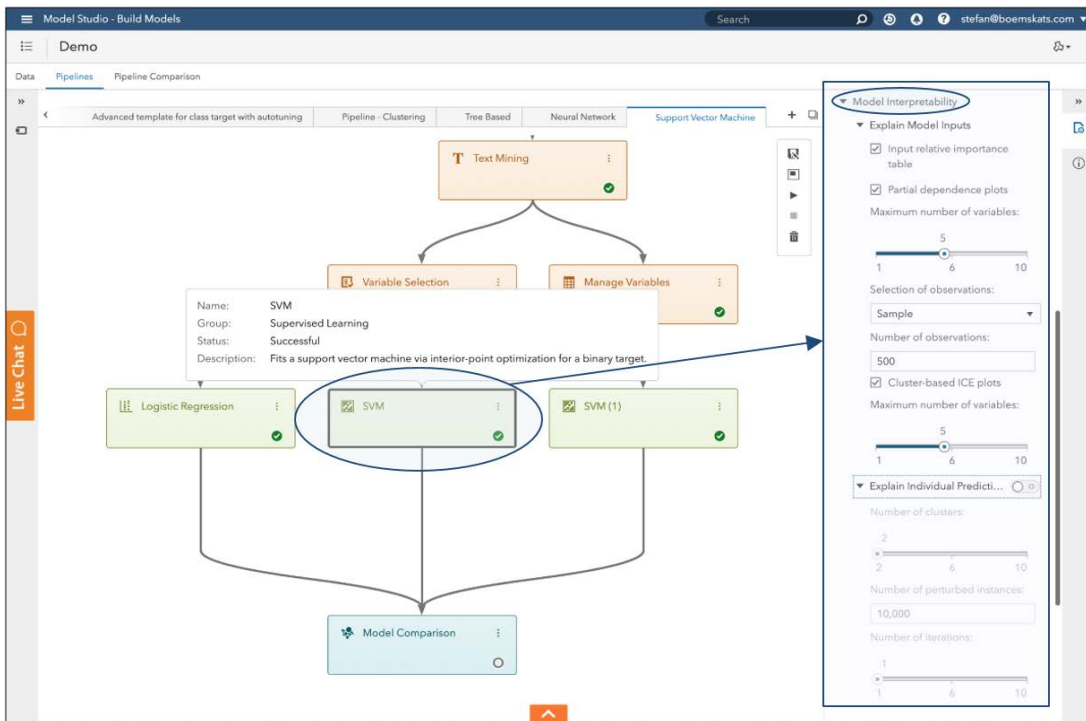
Display 11. Automated Model Interpretability Settings Pane in Model Studio

When you select the automated interpretability feature, it analyses the resulting predictions of the model. It creates several interactive model interpretability plots: The Partial Dependence (PD) plot, the Individual Conditional Expectation (ICE) plot, and the Local Interpretable Mode-Agnostic Explanations (LIME) plot. They visualize the relationship between the input variables and the target variable. Besides, the Input Relative Importance table presents you with the most important predictors in the model.
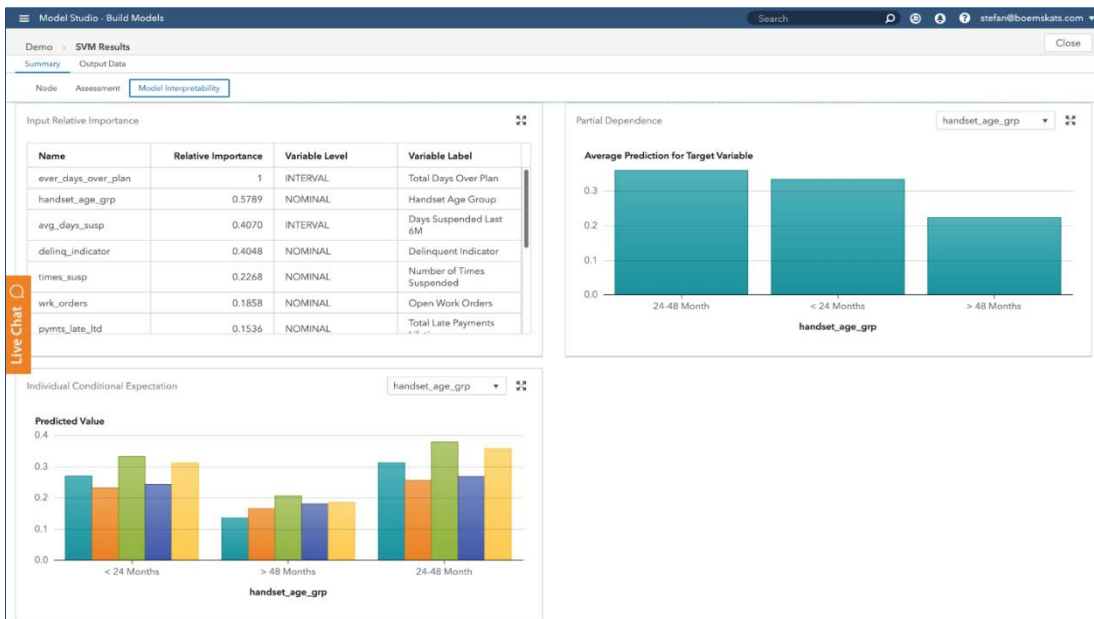

Display 12. Automated Model Interpretability Visualizations in Model Studio

In this way, SAS Viya transforms the "black box" of the SVM algorithm into an understandable and interpretable model that quickly equip your academic coursework with valuable business insights.

You can also use the model interpretability feature to compare different models generated by different Machine Learning algorithms. This opportunity additionally increases the transparency in choosing the best model of your coursework.

● Augmented analytics.

Moreover, SAS Viya for Learners applies natural language generation to present you the results of your analysis in easily understood terms. This feature boosts your data science knowledge absorption and enhances your learning curve.
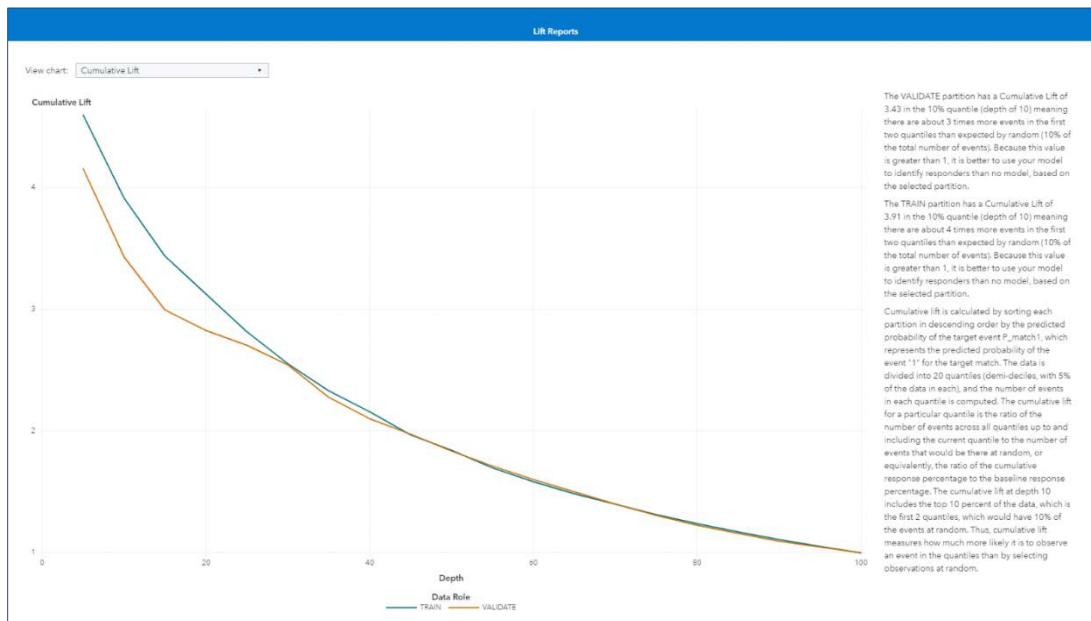

Display 13. A Cumulative Lift Chart with Natural Language Generated Explanation

● Seamless and secured collaboration in one place.

In SAS Viya for Learners, you can add and cooperate with multiple colleagues on group coursework. You can associate data with the coursework and share it with your team members. You can post updates. You can set project activity feeds to notify your course mates and team members of changes and updates (SAS Institute Inc., 2020). From SAS Data Studio, you can share and reuse existing data preparation plans (SAS Institute Inc., 2020). The Exchange feature of Model Studio enables you to quickly and easily create and share node or pipeline templates with your team. You can filter your search in the Exchange repository by type of Machine learning project (e.g., forecasting, text analytics) or kind of analytical task (e.g., preprocessing, modeling). Prebuilt model life cycle templates let you manage projects collaboratively.

Below, you can see the variety of prebuilt nodes and pipeline templates ready to be shared with your team and used in Model Studio:
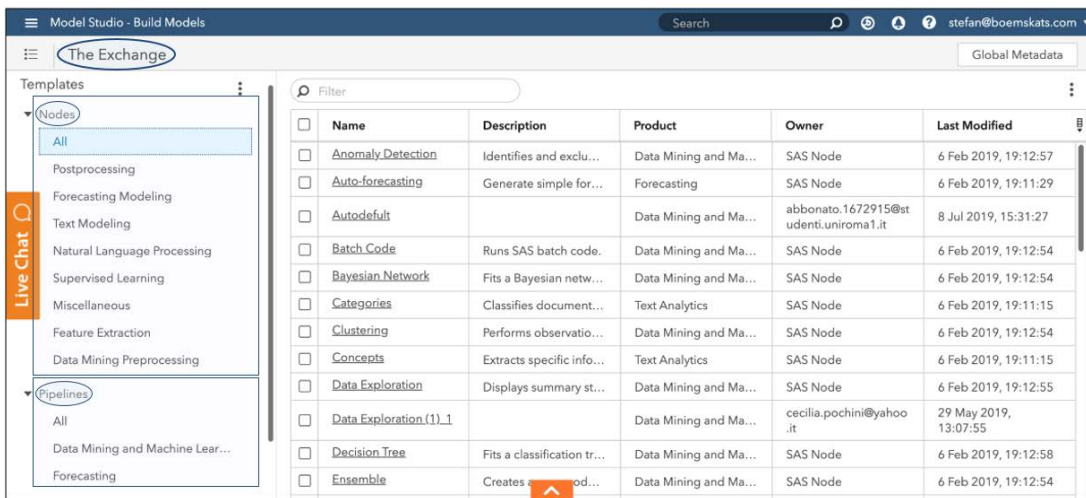
Display 14. The Exchange Tool in Model Studio

The distinctive collaboration features of SAS Viya for Learners enable you as an academic educator to design and assign coursework to interdisciplinary and diverse student teams. They can be more creative to reach better results in conditions closer to the real business world.

This cooperation boosts your students' productivity and enhances their data science skills as they can share knowledge and learn from each other. Moreover, they develop valuable soft skills for team working, communication, project management, time management, and leadership.

- Automated Model deployment to any environment.

  Once you have chosen the champion model of your analytics project, you can put it into production in any environment with just a few clicks in SAS Model Manager. SAS Viya automatically translates the model into a score code, which then can be run on a new dataset.
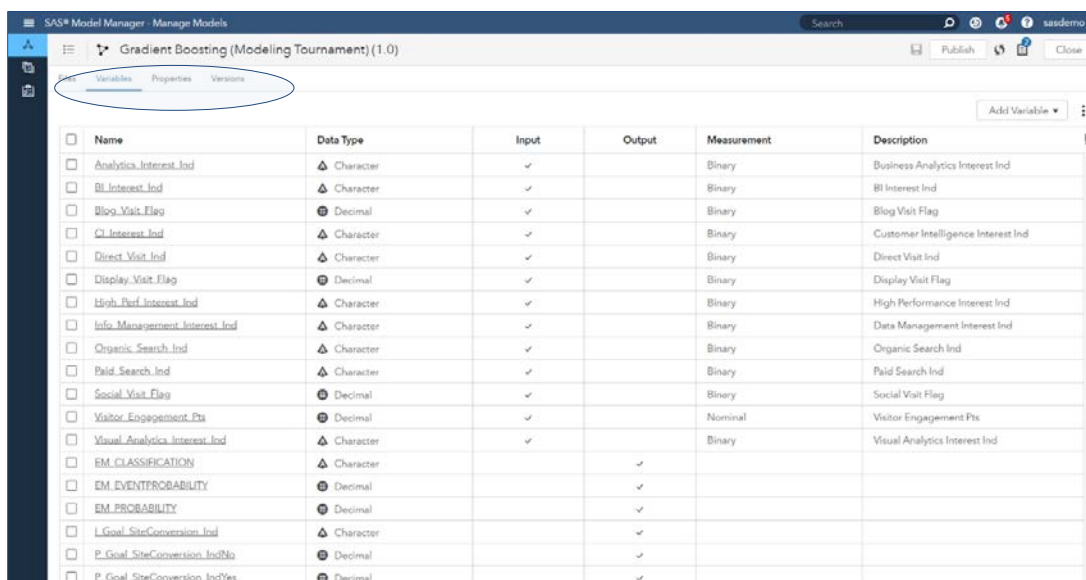
Display 15. SAS® Model Manager

- Convenient model and data governance accelerating the model life cycle and operationalization

  Moreover, within SAS Viya for Learners, you can seamlessly learn how to manage analytical models via a centralized, secure web-based repository. SAS Model Manager enables you to test and compare analytical models quickly. Within SAS Viya for Learners, you acquire skills to track models from creation, through usage, to retirement with a centralized, efficient, repeatable process for registering, validating, monitoring, and retraining models. You learn how to use automatic and comprehensive version control supported by SAS Viya.

  As a result, you get a solid understanding of how to provide visibility of your analytical lifecycle, ensuring complete traceability and analytics governance.

By comprehending all these features of SAS Viya, you equip yourself with highly valued skills on the job market for speed-up time to value delivery from your ML project.

RICH POOL OF LEARNING RESOURCES FOR EFFECTIVE DATA SCIENCE TEACHING

Training Courses for All Levels with Well-Structured High-Quality Content

SAS Viya for Learners provides you with many interactive courses. The extensive training pool empowers you to build a complete advanced analytics curriculum – from entry-level to doctorate – with a single software environment (SAS Institute, 2020).

You can choose and combine courses from a variety of learning paths, including Statistical Analysis, Predictive Analytics, Machine Learning, Text Analytics, Internet of Things, Visual Analytics, Programming, Forecasting and Econometrics, Data Scientist, and more (SAS Institute Inc., 2020).

SAS Viya for Learners presents you the concepts gradually from the simplest to the more advanced ones, building up your knowledge piece by piece. The platform offers knowledge in several different forms where each one complements the others, deepening your understanding gradually and with less stress:

- Engaging videos and comprehensive visualizations.

  Firstly, the concepts are presented and explained comprehensively in short, engaging, and informative videos. I was impressed that each word and sentence in the SAS videos seemed so carefully selected. There is no unnecessary information or overly long explanations. At the same time, interactive and dynamic visualizations in the videos enhance your comprehensiveness of the taught material. Moreover, SAS courses provide you with lots of demos of analytics procedures.

  I like training videos. Usually, with slides, the topics are explained just by few words in bullet points. While, in a video, you can hear a speaker explaining the concepts in a clear language and read the subtitles. You include more of your senses in the learning process. SAS videos are also complemented by a transcript that you can read additionally and thus enhance your knowledge absorption. The transcripts helped me a lot.

  SAS Institute also provides learners with a library of free "How To Tutorials" training videos to learn from as is displayed below (SAS Institute Inc., 2020):
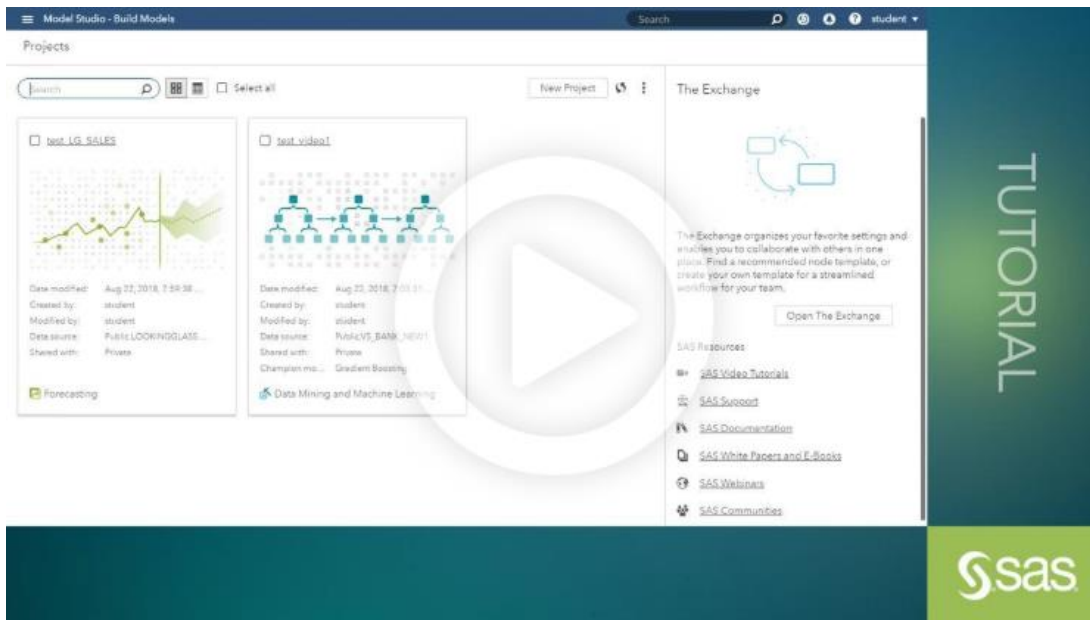
Display 16**. "Impute Missing Values in Model Studio" "How To Tutorial"** (SAS Institute Inc., 2020)

- Additional concise text summary on a topic.

  Many of the course videos are complemented by short informative reviews on some of the key subjects. Such text provides more details and lets you get an even better understanding of the concepts. For example, Table 1 represents you (on only two pages) with an excellent comparison of the most commonly used supervised Machine Learning algorithms.

  This comprehensive overview allows you to quickly understand how to choose the most appropriate Machine Learning algorithm for a specific analytical task, given the characteristics of your data.

  Such comprehensive teaching materials of the SAS Training courses help you to acquire a lot of relevant knowledge in addition to the rest of the university course materials.

- Plenty of practice materials with different levels of difficulty.

  After the presentation of each topic in the SAS training courses, there are several short practice tasks with different levels of difficulty. These exercises give you the chance to build your data science skills gradually, piece by piece, and without stress. Such an approach keeps you motivated throughout the whole learning journey.

- Quizzes with feedback at the end of each lesson.

  The quiz at the end of each lesson allows you to check your understanding of the taught topics. The feedback provides you with the right answers and comments on whether you are right or wrong. It refers you to the relevant teaching content to read again if needed. It made it easy for me to keep track of my progress and master what I learned.

MACHINE LEARNING QUICK REFERENCE: ALGORITHMS - 1

Algorithm Type	Common Usage	Suggested Usage	Suggested Scale	Interpretability	Common Concerns
Penalized Regression	• Supervised regression • Supervised classification	• Modeling linear or linearly separable phenomena • Manually specifying nonlinear and explicit interaction terms • Well suited for $N \ll p$	Small to large data sets	High	• Missing values • Outliers • Standardization • Parameter tuning
Naïve Bayes	Supervised classification	• Modeling linearly separable phenomena in large data sets • Well-suited for extremely large data sets where complex methods are intractable	Small to extremely large data sets	Moderate	• Strong linear independence assumption • Infrequent categorical levels
Decision Trees	• Supervised regression • Supervised classification	• Modeling nonlinear and nonlinearly separable phenomena in large, dirty data • Interactions considered automatically, but implicitly • Missing values and outliers in input variables handled automatically in many implementations • Decision tree ensembles, e.g. random forests and gradient boosting, can increase prediction accuracy and decrease overfitting, but also decrease scalability and interpretability	Medium to large data sets	Moderate	• Instability with small training data sets • Gradient boosting can be unstable with noise or outliers • Overfitting • Parameter tuning
k-Nearest Neighbors (kNN)	• Supervised regression • Supervised classification	• Modeling nonlinearly separable phenomena • Can be used to match the accuracy of more sophisticated techniques, but with fewer tuning parameters	Small to medium data sets	Low	• Missing values • Overfitting • Outliers • Standardization • Curse of dimensionality
Support Vector Machines (SVM)	• Supervised regression • Supervised classification • Anomaly detection	• Modeling linear or linearly separable phenomena by using linear kernels • Modeling nonlinear or nonlinearly separable phenomena by using nonlinear kernels • Anomaly detection with one-class SVM (OSVM)	• Small to large data sets for linear kernels • Small to medium data sets for nonlinear kernels	Low	• Missing values • Overfitting • Outliers • Standardization • Parameter tuning • Accuracy versus deep neural networks depends on choice of nonlinear kernel; Gaussian and polynomial often less accurate
Artificial Neural Networks (ANN)	• Supervised regression • Supervised classification • Unsupervised clustering • Unsupervised feature extraction • Anomaly detection	• Modeling nonlinear and nonlinearly separable phenomena • Deep neural networks (e.g. deep learning) are well suited for state-of-the-art pattern recognition in images, videos, and sound • All interactions considered in fully connected, multilayer topologies • Nonlinear feature extraction with autoencoder and restricted Boltzmann machine (RBM) networks • Anomaly detection with autoencoder networks • Clustering and visualization with self-organizing maps (SOMs)	• Usually small to medium data sets • Stochastic gradient descent (SGD) optimization drastically increases scalability	Low	• Missing values • Overfitting • Outliers • Standardization • Parameter tuning

https://github.com/sassoftware/enlighten-apply/tree/master/ML_tables   §sas

MACHINE LEARNING QUICK REFERENCE: ALGORITHMS - 2

Algorithm Type	Common Usage	Suggested Usage	Suggested Scale	Interpretability	Common Concerns
Association Rules	• Supervised rule building • Unsupervised rule building	Building sets of complex rules by using the co-occurrence of items or events in transactional data sets	Medium to large transactional data sets	Moderate	• Instability with small training data • Overfitting • Parameter tuning
k-Means	Unsupervised clustering	• Creating a known a priori number of spherical, disjoint, equally sized clusters • k-modes method can be used for categorical data • k-prototypes method can be used for mixed data	Small to large data sets	Moderate	• Missing values • Outliers • Standardization • Correct number of clusters is often unknown • Highly sensitive to initialization • Curse of dimensionality
Hierarchical Clustering	Unsupervised clustering	Creating a known a priori number of nonspherical, disjoint, or overlapping clusters of different sizes	Small data sets	Moderate	• Missing values • Standardization • Correct number of clusters is often unknown • Curse of dimensionality
Spectral Clustering	Unsupervised clustering	Creating a data-dependent number of arbitrarily-shaped, disjoint, or overlapping clusters of different sizes	Small data sets	Moderate	• Missing values • Standardization • Parameter tuning • Curse of dimensionality
Principal Components Analysis (PCA)	Unsupervised feature extraction	• Extracting a data-dependent number of linear, orthogonal features, where $N \gg p$ • Extracted features can be rotated to increase interpretability, but orthogonality is usually lost • Singular value decomposition (SVD) is often used instead of PCA on wide or sparse data • Sparse PCA can be used to create more interpretable features, but orthogonality is lost • Kernel PCA can be used to extract nonlinear features	• Small to large data sets for traditional PCA and SVD • Small to medium data sets for sparse PCA and kernel PCA	Generally low, but higher for sparse PCA or rotated solutions	• Missing values • Outliers
Nonnegative Matrix Factorization (NMF)	Unsupervised feature extraction	Extracting a known a priori number of interpretable, linear, oblique, nonnegative features	Small to large data sets	High	• Missing values • Outliers • Standardization • Correct number of features is often unknown • Presence of negative values
Random Projections	Unsupervised feature extraction	Extracting a data-dependent number of linear, uninterpretable, randomly-oriented features of equal importance	Medium to extremely large data sets	Low	Missing values
Factorization Machines	• Supervised regression and classification • Unsupervised feature extraction	• Extracting a known a priori number of uninterpretable, oblique features from sparse or transactional data sets • Can automatically account for variable interactions • Creating models from a large number of sparse features; can outperform SVM for sparse data	Medium to extremely large sparse or transactional data sets	Moderate	• Missing values • Outliers • Standardization • Correct number of features is often unknown • Less well suited for dense data

https://github.com/sassoftware/enlighten-apply/tree/master/ML_tables   §sas

Table 1. Machine Learning Quick Reference: Algorithms (SAS Institute Inc., 2016)

However, a theory is not everything. SAS has a much wider variety of materials designed to prepare you to become a real data scientist.

- Well-guided applied real-world business case studies

  Many SAS training courses provides you with the opportunity to start applying your new data science skills in a real scenario.

  Display 17 represents an introduction to a case study from the SAS Programming 1: Essentials course (SAS Institute Inc., 2020):

1.1 Case Study Introduction

In this case study, you solve a real-world business problem by applying concepts that you learned in the SAS Programming 1: Essentials course. Be aware that there are numerous solutions to this problem, and some can include concepts that are outside the scope of the SAS Programming 1 course.

Background Information

You are a SAS programmer with six months of experience who is in charge of creating basic reports and maintaining SAS programs. Recently, you completed a SAS Programming course, and your supervisor gives you your first SAS programming project.

Business Problem

Your first project is to prepare and analyze Transportation Security Administration (TSA) Airport Claims data from 2002 through 2017. The TSA is an agency of the United States Department of Homeland Security that has authority over the security of the traveling public. A claim is filed if you are injured or your property is lost or damaged during the screening process at an airport.

To complete your project, you follow your supervisor's requirements, which are in Section 1.3 of this document. Here is what you need to do:

- Prepare the data.
- Create one PDF report that analyzes the overall data as well as the data for a dynamically specified state.

Display 17. Case Study Introduction from the "SAS Programming 1: Essentials" Training Course

  I feel I learned the most from this part. SAS helped me implement the business task with guidance and hints and taught me how to approach a real-world problem.

  I was able to finish the case study in only one day. The most significant benefit was that the case study taught me to think critically. I had to choose the most appropriate analytical procedures and options to solve a real-world, complex task. It helped me to develop my problem-solving skills as well.

SAS helps you to build your data science skills. Starting from the most straightforward concepts and building piece by piece your skills until you can solve a real-world business challenge.

SAS Documentation, Examples and Papers

After you finish a training course and you have solid foundations on the taught topics, you can explore even more analytical concepts. SAS has the right materials for the next level of your studies:

- SAS Documentation.

  The training videos provide references to advanced SAS documentation (SAS Institute Inc., 2020), and guidance for conducting analytics in specific business areas, such as

healthcare analytics or financial fraud detection. Using this, you can learn in more detail how to put the wide variety of features of the analytical techniques into practice.

- SAS Examples.

  Within SAS Support, you can find examples for implementing a variety of analytical tasks. You can browse them by topic - for example, for data preparation (SAS Institute Inc., 2020) or statistical procedures (SAS Institutre Inc., 2020). Also, the SAS knowledge base provides you with lots of practical examples from real-life industries with helpful explanations on how to implement different analytical techniques, specifically within the SAS environment (SAS Institute Inc., 2020). In this way, as a student, you boost faster your analytical skills tailored to specific business domains. SAS supports you to get better prepared and be more employable for the job market.

  You can see in Display 18 below a SAS example that shows you how a petroleum company which uses a turbine to heat water into steam can analyze the turbine performance by creating a box plot:

The input data set *Turbine* is specified with the DATA= option in the PROC BOXPLOT statement. The PLOT statement requests a box-and-whiskers plot for each group of data. After the keyword PLOT, you specify the analysis variable (in this case, *KWatts*), followed by an asterisk and the group variable (*Day*). The ODS GRAPHICS OFF statement specified before the PROC BOXPLOT statement disables ODS Graphics, so the box plot is produced using traditional graphics. The box plot is shown in Figure 29.1.

Figure 29.1: Box Plot for Power Output Data
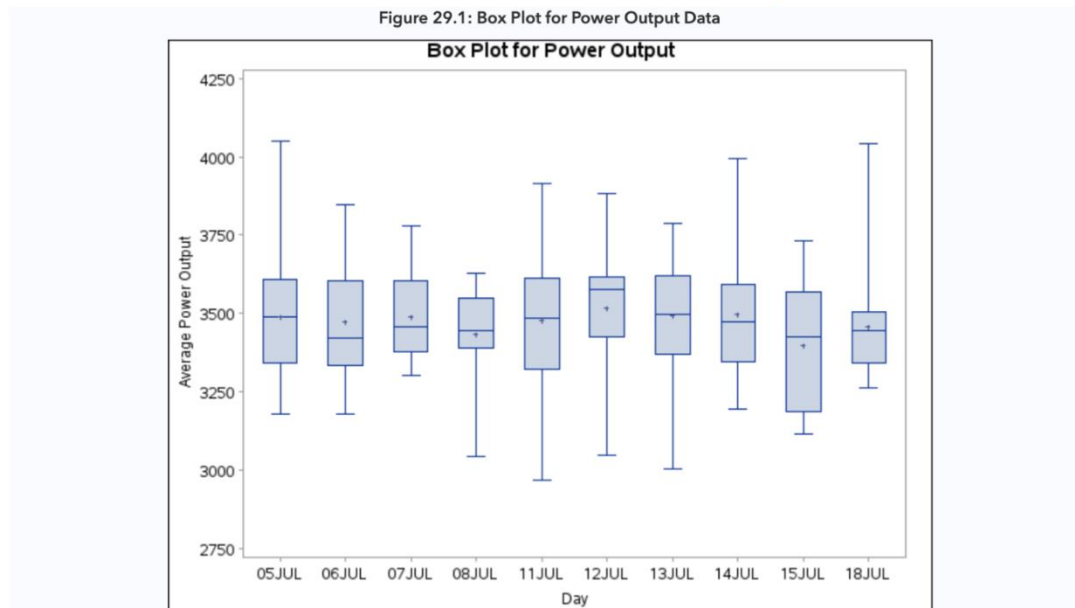
The box plot displayed in Figure 29.1 represents summary statistics for the analysis variable *KWatts*. Each of the 10 box-and-whiskers plots describes the variable *KWatts* for a particular day. The plot elements and the statistics they represent are as follows:

- The length of the box represents the interquartile range (the distance between the 25th and 75th percentiles).
- The symbol in the box interior represents the group mean.
- The horizontal line in the box interior represents the group median.
- The vertical lines (called *whiskers*) issuing from the box extend to the group minimum and maximum values.

Display 18. Box Plot and Explanations from a SAS Example

- SAS papers.

  Furthermore, you can read plenty of SAS papers (SAS Institute Inc., 2020). They teach you how to integrate and complete different tasks in the unified SAS environment to achieve a goal as part of the whole analytics cycle.

High-Quality Personalized Learning Support

One of the main advantages of SAS Viya for Learners that enhances the academic teaching of data science topics is the broad learning support it provides.

In many cases, as a university professor, you have a lot of students or high research workload. Therefore, you don't have enough time to provide a quick and personalized answer to every single student.

SAS' extensive team of highly qualified educational experts is dedicated to support you and quickly answer your students' questions via various communication channels. They quickly identify a student's level of understanding of the concepts and guide them in the right direction.

- Online chat.

    If you as a student have a specific question on a topic from a SAS training course and you feel you can't move to the next section without understanding the previous one, you can instantly ask a question by clicking on the "chat" button. You receive a comprehensive and concise answer with detailed explanations and guidance in no more than 24 hours. Often it takes only several hours to get them. The quick answers help you to understand the concepts faster.

    For example, while taking the 'Machine Learning using SAS Viya' course, I decided to click on the 'chat' button to ask a question about one machine learning algorithm. I received a comprehensive and detailed email explanation only a few hours later. I fully understood the advanced concept in just five minutes.

    Such learning support keeps you on track and motivated to continue learning advanced analytics topics.

- Learning communities.

    The online SAS communities help you to share data science knowledge with peers not only from your university course but from around the world (SAS Institute Inc., 2020). In this way, you can easily browse from existing Q&A or ask a new question to get fast, helpful answers which deliver to you many perspectives of views reflecting more broader backgrounds. SAS users can also share a well-organized sample SAS or open-source code (Hemedinger, 2020). In this way, SAS communities enhance your students' analytics skills at scale.

- "Ask the Expert" webinars.

    Universities sometimes organize guest lectures, usually once or twice per academic semester. SAS provides your students with the opportunity to attend many more free online live sessions throughout the whole academic year (SAS Institute Inc., 2020). Students can ask and get answers to specific questions from well-established industry experts from all around the globe. Moreover, you can view previously recorded sessions on-demand to get in-depth information about a particular topic. This flexibility enables you to boost much more your understanding of many data science topics much faster.

- Personalized video tutorials.

    Furthermore, the SAS educational experts even create short personalized video tutorials to answer questions students ask them.

    For example, when practicing my coding as part of one of the advanced SAS modules, I saw an error in the program log, which I was sure should not be there. Therefore, I emailed support@sas.com explaining my problem. A SAS Education expert made a 10

min YouTube video tutorial for me to help me fully understand the answer to my question (Zender, 2019).

This approach makes the learning support even more interactive, which helps you to comprehend quickly and more in-depth a specific concept. It inspires you and motivates you even more. In this way, the timely, interactive, and comprehensive SAS Learning Support prevents any bottlenecks in your studying and teaching workflow, making the learning process of data science less frictionless and more seamless.

Free SAS Viya for Learners Educators' Resources

In addition to the variety of student learning tools, SAS Viya provides a variety of resources to support academic educators to create effective data science curriculums and enhance the teaching process. Here are some of them:

- SAS Viya for Learners Private Forum for Educators.

  This professional community enables you to share your experience and thoughts as a data science educator with peers from around the world (SAS Insititute Inc., 2020). You can easily and quickly exchange and discuss ideas for enhancements in the teaching process. By accumulating innovative and creative thinking from all corners of the globe, you can work together to build the best advance-analytics academic curriculum and delivery platform.

- Free high-quality teaching materials and books for educators.

  SAS can provide you as an academic educator with free high-quality teaching books (SAS Institute Inc., 2020) and additional training material like Microsoft PowerPoint slides, course notes, and data sets (SAS Institute Inc., 2020). They are developed and updated by industry-leading experts. In this way, you save time while preparing for teaching. Moreover, you can customize the teaching materials to integrate them into your creative curriculum.

- Free Professor Workshops.

  During these workshops, you acquire the skills to teach students on a variety of topics, including advanced business analytics, data mining, machine learning, and more (SAS Institute Inc., 2020). Besides, these seminars help you to strengthen your analytical skills and to keep up with the latest developments in data science.

HOW SAS VIYA FOR LEARNERS COULD BE EVEN BETTER WITH DATA-DRIVEN APPS

SAS Viya supports a REST API, designed to allow data scientists to embed smart insights into mobile and web apps. These capabilities of the analytics platform provide the potential **further to enhance student's development as a well**-rounded data scientist.

Boemska AppFactory democratizes the development of data apps by helping data scientists collaborate with app developers to build apps and experiences that are greater than the sum of their parts. It could, therefore, be included in academic curriculums to reach the next level of data science education, encourage interdisciplinary collaboration between data scientists and computer scientists and empower students with different skills to learn how to complete the last mile of analytics at scale.

Here is a quick overview of SAS Viya and Boemska infrastructure:

Figure 3. Overview of SAS® Viya® and Boemska AppFactory Infrastructure

To the left, we have the data sources. Next, data is imported and analyzed in SAS Viya. The generated insights are input in Boemska AppFactory.

One of the main benefits of Boemska AppFactory is that it provides an integrated development experience. It is simplifying and accelerating the collaboration between the ML model development and application development teams.

AppFactory facilitates a way of working that lets each team continue using their preferred environments, frameworks, toolsets, and programming languages without asking them to change their usual workflow.

Minimizing the dependencies between teams increases developer productivity on both ends dramatically and makes the app development and analytical insights' deployment faster and more frictionless.

It feeds data directly to JavaScript runtimes and can be easily plugged into pre-built components, such as interactive charts in JavaScript, thus delivering superior and unique user experience.

Through this approach, AppFactory enables quick delivery of tailored business-led and data-centric applications. The apps can embed insights into the day-to-day tasks of many non-technical business users and stakeholders faster. Thus, enabling for better pro-active decision making and innovation, which leads to increased competitiveness and business results.

By learning Boemska AppFactory, students could gain skills to complete the last mile of analytics quickly, efficiently, and at a scale.

CONCLUSION

SAS Viya for Learners offers many advanced tools for studying data science. The fast cloud suite especially helpful to overcome common obstacles in academic teaching that hinder novice learners, such as: extensive and challenging to learn textbooks, lack of alternative concise learning sources, outdated or complicated teaching software.

The advantages of SAS Viya for Learners over IBM SPSS Modeler and R makes it much easier for novice students to comprehend machine learning concepts and appreciate the entire analytics life cycle. Model Studio and other solutions of SAS Viya enables students' collaboration, combining well-structured visual analytics workflow with SAS and open-source code, parameters that are easy to set up, an added layer of interpretability, automated hyperparameter tuning, assessment visualizations, best model selection and deployment, and convenient model governance.

Also, SAS Viya for learners provides students with a rich pool of learning resources, such as: interactive courses in statistics, programming, and advanced analytics, videos, summaries, examples, practices, case studies, papers, documentation, online support from educational experts, webinars, and communities. They enable novice learners to advance in their studies of the taught concepts more smoothly, quickly, and efficiently. In addition to the rich pool of student learning tools, SAS Viya provides a variety of resources to support academic educators, such as a private forum for educators, free high-quality teaching materials and books, and professor workshops. These tools empower academics to create more effective and innovative data science curriculums and boost the teaching process.

Moreover, SAS Viya infrastructure enables data scientists to embed smart insights into mobile and web apps. The capabilities of the analytics platform provide the potential further to enhance student's development as a well-rounded data scientist. Therefore, innovative tools like Boemska AppFactory, which democratize the development of data apps, could be included in academic curriculums to reach the next level of data science education. This approach will encourage interdisciplinary collaboration between data scientists and computer scientists. It will empower students with a variety of skills to complete the last mile of analytics at scale.

By including SAS Viya for Learners in academic curriculums, I firmly believe that educators could teach more creatively to speed-up students' deep data science comprehension, providing better-prepared professionals to bridge the gap in data science skills.

REFERENCES

Bathaee, Y., 2018. The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology,* 31(2), pp. 889-938.

Berman, J., 2019. *IDC: AI Can Significantly Help Organizations with Analytics, Business Intelligence.* [Online]
Available at: https://www.mesalliance.org/2019/09/27/idc-ai-can-significantly-help-organizations-with-analytics-business-intelligence/
[Accessed 8 January 2020].

Burning Glass Technologies, 2017. *The Quant Crunch: How The Demand For Data Science Skills Is Disrupting The Job Market,* BOSTON, MA 02110 USA: Burning Glass Technologies, Business-Higher Education Forum (BHEF) and IBM.

Columbus, L., 2017. *IBM Predicts Demand For Data Scientists Will Soar 28% By 2020.* [Online]
Available at: https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#256df6017e3b
[Accessed 15 October 2019].

Committee, G. C. R. A. R., 2016. *Guidelines for Assessment and Instruction in Statistics Education College Report 2016,* Alexandria, Virginia, USA: American Statistical Association .

DataFlair, 2019. *Pros and Cons of R Programming Language – Unveil the Essential Aspects!.* [Online]
Available at: https://data-flair.training/blogs/pros-and-cons-of-r-programming-

language/?utm_campaign=News&utm_medium=Community&utm_source=DataCamp.com
[Accessed 26 February 2020].

Gillespie, C. & Lovelace, R., 2016. 9 Efficient collaboration. In: *Efficient R programming.* s.l.:O'Reilly Media, Inc..

Harraway, J. A., 2012. Learning Statistics Using Motivational Videos, Real Data and Free Software. *Technology Innovations in Statistics Education,* 6(1).

Hemedinger, C., 2020. *How to add SAS syntax to your post - SAS Support Communities.* [Online]
Available at: https://communities.sas.com/t5/Getting-Started/How-to-add-SAS-syntax-to-your-post/ta-p/224394
[Accessed 25 February 2020].

Hinchcliffe, A., 2019. *SAS Analytics Experience: Removing hurdles.* [Online]
Available at: https://www.itweb.co.za/content/LPwQ57lyoLRMNgkj
[Accessed 8 January 2020].

IBM, 2020. *SPSS Modeler - Overview - United Kingdom | IBM.* [Online]
Available at: https://www.ibm.com/uk-en/products/spss-modeler
[Accessed 26 February 2020].

Khan, M. M. A., 2019. *Now, educators can access SAS' Viya software for free.* [Online]
Available at: https://www.thehindubusinessline.com/info-tech/now-educators-can-access-sas-viya-software-for-free/article27003757.ece
[Accessed 24 February 2020].

Molnar, C., 2019. *Introduction to iml: Interpretable Machine Learning in R.* [Online]
Available at: https://cran.r-project.org/web/packages/iml/vignettes/intro.html
[Accessed 27 February 2020].

Molnar, C., 2020. Chapter 5 Model-Agnostic Methods. In: C. Molnar, ed. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable..* s.l.:Molnar, Christoph, p. 251.

PwC, 2017. *Data science and analytics job market predictions: PwC.* [Online]
Available at: https://www.pwc.com/us/en/library/data-science-and-analytics.html
[Accessed 27 February 2020].

Reddy, K. J., Menon, K. R. & AnjanaThattil, 2018. Academic Stress and its Sources Among University Students. *Biomed Pharmacol,* 11(1).

Robert Walters Group, 2018. *Solving the UK Skills Shortage - Technology.* [Online]
Available at: https://www.robertwalters.co.uk/solving-the-uk-skills-shortage/technology-research.html
[Accessed 15 October 2019].

Ross, N., 2013. *Don't R alone! A guide to tools for collaboration with R.* [Online]
Available at: https://www.noamross.net/archives/2013-01-07-collaborating-with-r/
[Accessed 26 February 2020].

SAS Insititute Inc., 2020. *SAS® Viya® for Learners Private Educator Forum - SAS Support Communities.* [Online]
Available at: https://communities.sas.com/t5/SAS-Viya-for-Learners-Private/gp-p/V4L
[Accessed 25 February 2020].

SAS Institute Inc., 2019. *Machine Learning Using SAS® Viya®.* [Online]
Available at: https://support.sas.com/edu/schedules.html?ctry=us&crs=CPML83
[Accessed 27 February 2020].

SAS Institute Inc., 2020. *Ask the Expert - Webinarss | SAS.* [Online]
Available at: https://www.sas.com/en_us/learn/ask-the-expert-webinars.html
[Accessed 25 February 2020].

SAS Institute Inc., 2020. *Free SAS Academic Programs Teaching Materials.* [Online]
Available at: https://www.sas.com/en_gb/learn/academic-programs/resources/free-teaching-

materials.html

[Accessed 25 February 2020].

SAS Institute Inc., 2020. *Free SAS Books for Educators.* [Online]

Available at: https://www.sas.com/en_gb/learn/academic-programs/resources/free-book-evaluation-copies.html

[Accessed 25 February 2020].

SAS Institute Inc., 2020. *Free SAS Professor Workshops.* [Online]

Available at: https://www.sas.com/en_gb/learn/academic-programs/resources/free-professor-workshops.html

[Accessed 25 February 2020].

SAS Institute Inc., 2020. *Imputing Missing Values in Model Studio - Model Studio - SAS Video Portal.* [Online]

Available at: https://video.sas.com/detail/videos/model-studio/video/5833452377001/imputing-missing-values-in-model-studio?autoStart=true

[Accessed 25 February 2020].

SAS Institute Inc., 2020. *Knowledge Base | SAS Support.* [Online]

Available at: https://support.sas.com/en/knowledge-base.html

[Accessed 25 February 2020].

SAS Institute Inc., 2020. *Learn SAS - SAS Support Communities.* [Online]

Available at: https://communities.sas.com/t5/Learn-SAS/ct-p/learn

[Accessed 25 February 2020].

SAS Institute Inc., 2020. *Model Studio Customer Documentation Page.* [Online]

Available at: https://support.sas.com/documentation/prod-p/model-studio/index.html

[Accessed 27 February 2020].

SAS Institute Inc., 2020. *SAS Documentation | SAS Support.* [Online]

Available at: https://support.sas.com/en/documentation.html

[Accessed 25 February 2020].

SAS Institute Inc., 2020. *SAS Educaton | Training | Learning Paths.* [Online]

Available at: http://www.sas.com/offices/europe/uk/education/paths/

[Accessed 25 Februaryq 2020].

SAS Institute Inc., 2020. *SAS Help Center: SAS® Data Explorer 2.5: User's Guide.* [Online]

Available at:
https://documentation.sas.com/?cdcId=dprepcdc&cdcVersion=2.5&docsetId=datahub&docsetTarget=titlepage.htm&locale=en

[Accessed 27 February 2020].

SAS Institute Inc., 2020. *SAS Help Center: SAS® Data Studio 2.5: User's Guide.* [Online]

Available at:
https://documentation.sas.com/?cdcId=dprepcdc&cdcVersion=2.5&docsetId=datastudioadv&docsetTarget=titlepage.htm&locale=en

[Accessed 27 February 2020].

SAS Institute Inc., 2020. *SAS Model Manager | SAS Support.* [Online]

Available at: https://support.sas.com/en/software/model-manager-support.html#documentation

[Accessed 27 February 2020].

SAS Institute Inc., 2020. *SAS Training in United Kingdom - SAS Programming 1: Essentials.* [Online]

Available at: https://support.sas.com/edu/schedules.html?ctry=gb&crs=PROG1#s1=1

[Accessed 25 February 2020].

SAS Institute Inc., 2020. *SAS Video Portal.* [Online]

Available at: https://video.sas.com/category/videos/how-to-tutorials

[Accessed 25 February 2019].

SAS Institute Inc., 2020. *SAS® Data Preparation: Tutorials and Examples.* [Online]
Available at:
https://documentation.sas.com/api/docsets/dprepex/2.1/content/dprepex.pdf?locale=en#nameddest=titlepage
[Accessed 25 February 2020].

SAS Institute Inc., 2020. *SAS® Data Preparation.* [Online]
Available at: https://www.sas.com/en_us/software/data-preparation.html
[Accessed 27 February 2020].

SAS Institute Inc., 2020. *SAS® Visual Analytics.* [Online]
Available at: https://www.sas.com/en_us/software/visual-analytics.html
[Accessed 27 February 2020].

SAS Institute Inc., 2020. *Technical Papers | SAS Support.* [Online]
Available at: https://support.sas.com/en/technical-papers.html
[Accessed 25 February 2020].

SAS Institute, 2020. *SAS Viya for Learners.* [Online]
Available at: https://www.sas.com/en_gb/software/viya-for-learners.html#for-students
[Accessed 24 February 2020].

SAS Institutre Inc., 2020. *SAS/STAT® Examples.* [Online]
Available at: https://support.sas.com/rnd/app/examples/STATexamples.html
[Accessed 25 February 2020].

Smith, T., 2020. *Common Problems With Open Source.* [Online]
Available at: https://dzone.com/articles/common-problems-with-open-source
[Accessed 26 February 2020].

Stoyanov, S., 2019. *Boosting your Machine Learning productivity with SAS Viya | Data Science Central.* [Online]
Available at: https://www.datasciencecentral.com/profiles/blogs/boosting-your-machine-learning-productivity-with-sas-viya
[Accessed 27 February 2020].

Stoyanov, S., 2019. *Boosting your Machine Learning Productivity with SAS® Viya®.* [Online]
Available at: https://www.linkedin.com/pulse/boosting-your-machine-learning-productivity-sas-viya-stefan-/
[Accessed 27 February 2020].

Stoyanov, S., 2019. *Professional Development Review,* Guildford, UK: University of Surrey.

The R Foundation, 2020. *CRAN Mirrors.* [Online]
Available at: https://cran.r-project.org/mirrors.html
[Accessed 26 February 2020].

The R Foundation, 2020. *R: What is R?.* [Online]
Available at: https://www.r-project.org/about.html
[Accessed 26 February 2020].

The Royal Society, 2019. *Growing demand for data science leaves Britain vulnerable to skills shortages.* [Online]
Available at: https://royalsociety.org/news/2019/05/data-science-skills-shortages/
[Accessed 27 February 2020].

University of Surrey, 2018. *Econometrics 1 - 2018/9 - University of Surrey.* [Online]
Available at: https://catalogue.surrey.ac.uk/2018-9/module/ECOM042
[Accessed 27 February 2020].

University of Surrey, 2018. *MACHINE LEARNING & VISUALISATIONS - 2018/9.* [Online]
Available at: https://catalogue.surrey.ac.uk/2018-9/module/MANM354
[Accessed 26 February 2020].

Zender, C., 2019. *Libname and outpath EPG194 | SAS Edu Customer Support.* [Online]
Available at: https://www.youtube.com/watch?v=PzQordBXc10&feature=youtu.be&hd=1
[Accessed 25 February 2020].

ACKNOWLEDGMENTS

RECOMMENDED READING

- *The Quant Crunch: How the Demand for Data Science Skills is Disrupting the Job Market*
- *The Dynamics of Data Science Skills Royal Society Report*
- *Exploring SAS® Viya® free e-books series*
- *SAS® Viya®: The R Perspective*
- *SAS® Viya®: The Python Perspective*
- *SAS® Visual Data Mining and Machine Learning 8.5: User's Guide*
- *SAS Machine Learning Blogs*

CONTACT INFORMATION

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

Stefan Dimitrov Stoyanov
MSc Business Analytics student
University of Surrey
stefdstoyanov@gmail.com
https://www.linkedin.com/in/stefan-dimitrov-stoyanov/