

Story Telling with SAS® Visual Statistics

Jim Box and Laura Watson, SAS Institute Inc.

ABSTRACT

SAS® Visual Statistics allows you to combine creative data visualizations with advanced modeling techniques to create compelling stories to solve data problems. In this presentation, we'll combine dashboards, visualizations, and statistical analyses to find the root causes for a medical procedure failure.

INTRODUCTION

SAS® Visual Statistics on the Viya platform is an application that combines visualization tools, variable creation tools, graphical controls, automated analytics, and statistical modeling methods. It is a powerful suite that allows the user to tell stories about data problems and solutions.

Telling a good story about a data challenge follows this general flow:

- Describe the problem
- Show the Solution
- Show the path to the results
 - Explore the Data
 - Describe the Analysis
- Conclusions & Next Steps

In this paper, we'll follow this flow.

DESCRIBE THE PROBLEM

In moving through the broader topic of creating a story with data using SAS Visual Statistics, we will consider in tandem a meaningful, real-world example to help illustrate the process. The example dataset includes cardiovascular patients from Cambridge, England, treated with arterial stents – a certain percentage of which experienced stent failure.

The first step in searching for a story within the data is to define the instigator, or reason for which data must be analyzed. Similarly, we define the outcome of interest and possible underlying factors, then organize these variables into categories, such as demographic, exposure, and outcome variables. Through data exploration, we aim to identify relationships existing between the variables in our model and whether they influence the outcome of interest.

For our situation, ~27% of patients experienced stent failure, the outcome of interest. Underlying factors include demographic, procedural, and stent material variables. Using this dataset, we seek to identify risk factors for stent failure and understand relationships between our included variables and stent failure.

To explain the problem, we'll want to use a combination on graphics, text, and numbers. Visual Statistics lets you put different items on your pallet and allow you to move them around as necessary. Display 1 shows how we illustrate the procedure, provide details on the problem, and give sample size and outcome rate variables of interest.

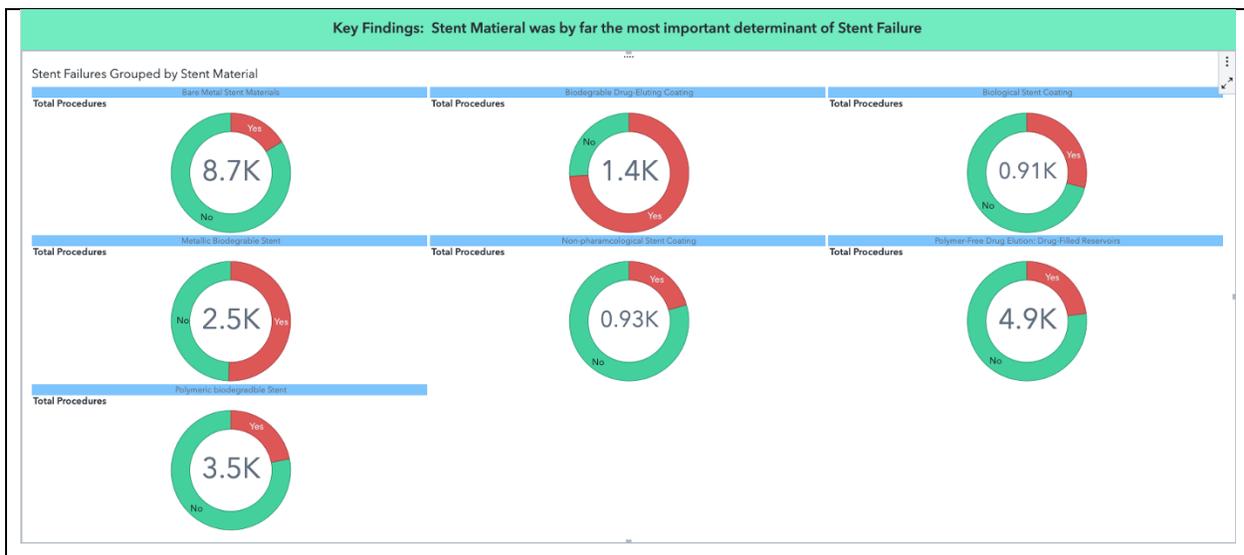


Display 1. Describe the problem we are trying to solve and identify the data available

SHOW THE SOLUTION

Typically, when addressing a data driven problem, the audience is more interested in the results than the methodology. While the analytical methods may be of great interest to those of us making these sorts of reports, the businessperson you are creating this for will be far more interested in the results. Don't make them wait – bring the most interesting results right up front. This will likely pique their interest to see how you got there, but they may only care about the solutions. That's OK, think of them as your customer and give them what they need. The rest of the report is the justification for your results and is also important, but don't make your audience wait to see the answer.

In our case, as Display 2 shows, the main driver of the stent failure rate among the studied procedures is the stent material itself. Procedures done with two of those materials had a demonstrably higher failure rate than those done with the others.



Display 2. Stent Material was by far the most important determinant of procedure failure.

EXPLORE THE DATA

SAS Visual Statistics gives you many different visualization objects to choose from. Display 3 shows the types you can choose from, including geographical objects to make map charts.

<p>▼ Graphs</p> <ul style="list-style-type: none">Bar chartBox plotBubble change plotBubble plotButterfly chartComparative time series plotCorrelation matrixDot plotDual axis bar chartDual axis bar-line chartDual axis line chartDual axis time series plotGaugeHeat mapHistogram	<p># Key value</p> <ul style="list-style-type: none">Line chartNeedle plotNumeric series plotParallel coordinates plotPie chartScatter plotSchedule chartStep plotTargeted bar chartTime series plotTreemapVector plotWaterfall chartWord cloud	<p>▼ Geo maps</p> <ul style="list-style-type: none">Geo bubbleGeo clusterGeo contourGeo coordinateGeo lineGeo line-coordinateGeo networkGeo pieGeo regionGeo region-coordinate
--	--	---

Display 3. Graph and Geo-map objects.

After sufficiently outlining the problem to be addressed, we explore our dataset to gain a sense of its composition. Begin by simply looking at your dataset in a crosstab (excel) format (Image 2). Ensure that there aren't missing data or easily noticeable documentation errors. Upon finding a mistake, navigate to the Develop SAS Code feature in SAS Visual Analytics. Use SAS Studio to properly format the data and correct any errors, then re-upload the data to the SAS Visual Analytics platform.

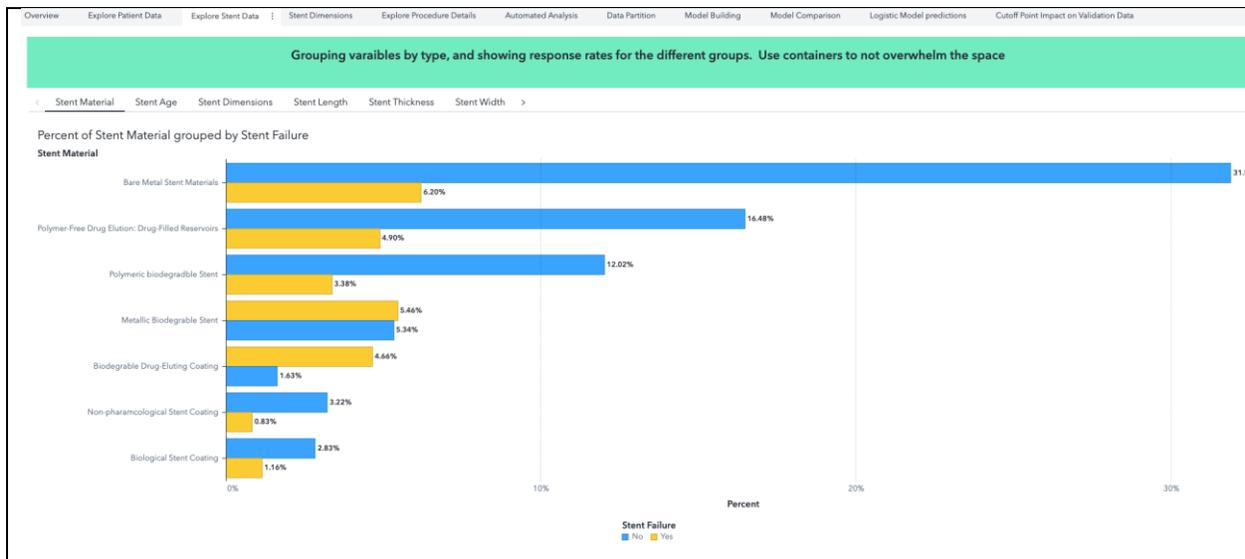
Next, generate visualizations that represent the distributions for the demographic and exposure variables in the data. Below is a list of recommended visualizations for that purpose:

- Bar Chart: represents categorical data with rectangles proportional to their values (Display 4)
- Box Plot: demonstrates the spread of continuous data, noting mean and quartile values
- Dot plot: uses dots to express the relationship between variables as coordinates
- Gauges: expresses a single value, such as a percentage or count, in context of a target range for that value
- Heat Map: represents a measure value, expressed by category, on a two-dimensional color scale
- Pie Chart: a circular graphic representing the proportions of each included category
- Tree Map: displays categories as rectangles, where the size and color of each rectangle represent a measure value for its respective category

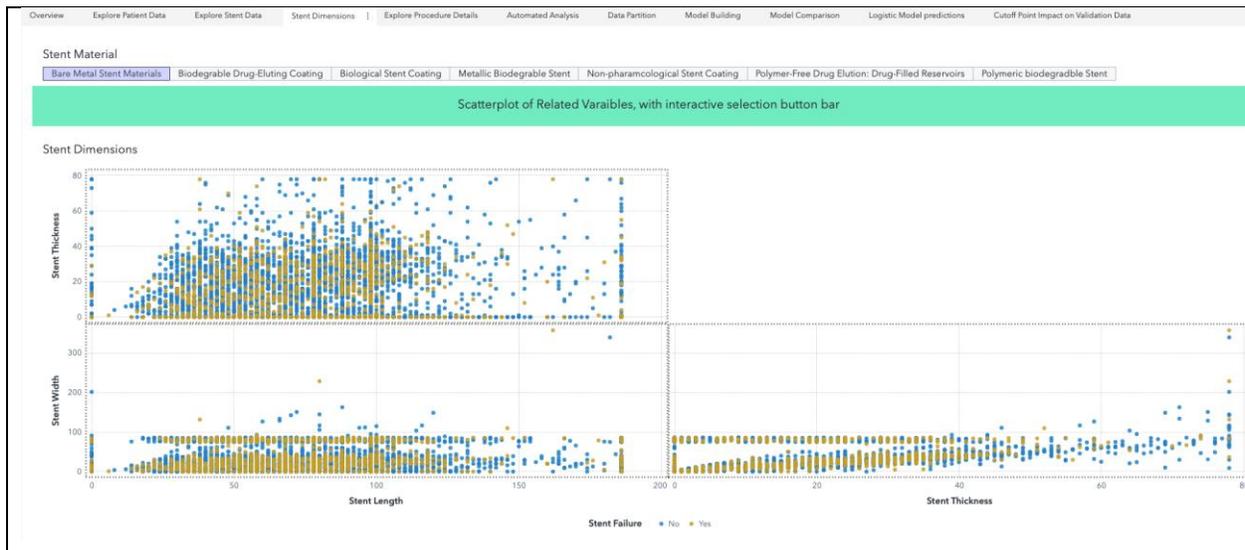
Finally, determine if the outcome of interest varies according to specific demographic or exposure variables and create the appropriate visualizations. Below is a list of tools useful for creating such visualizations:

- Button Bar: segments data shown on the page by category. Select a category by clicking the desired button. Useful for categories with few tiers. (Display 5)

- Drop Down Lists: segments data shown on the page by category. Select category by choosing from a drop-down list. Useful for categories with many tiers.
- Group-By Feature: segment data by category on a single graph. Useful for comparing distributions of measures relating to binary outcomes.
- Lattices: displays separate graphs for each value of a category. Useful for comparing non-overlaid graphs
- Stacking Container: Groups graphs together on the same page by stacking on top of another, navigable by tabs. Useful for organizing graphs such that all graphs on one page pertain to the same question (Display 4)



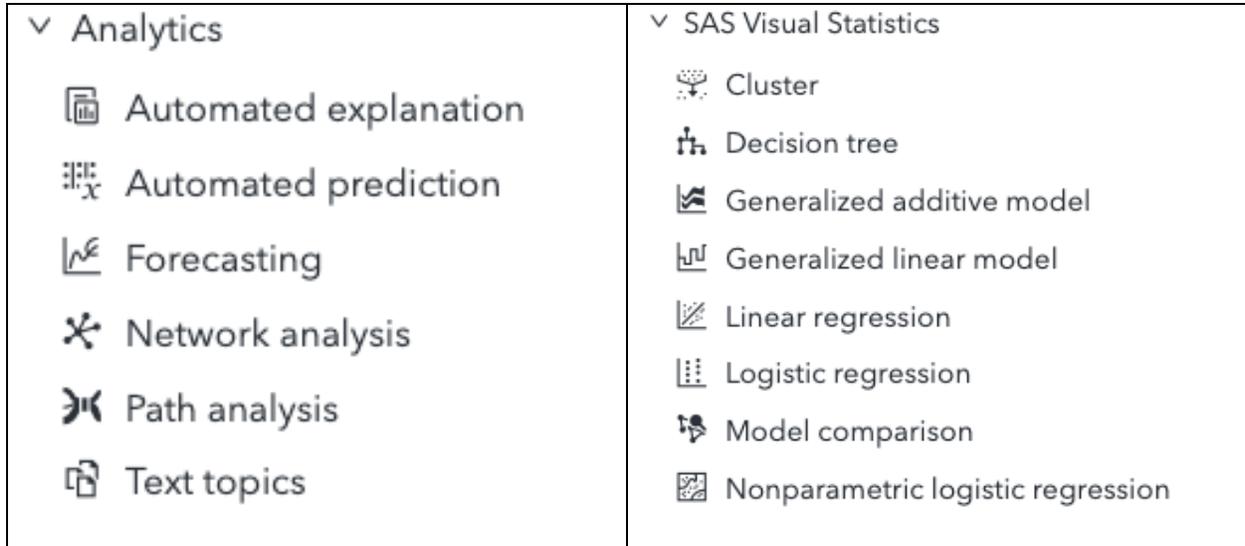
Display 4. Stacking container with multiple bar plots of related variables.



Display 5. Scatterplot of multiple variables, with a button bar control for filtering.

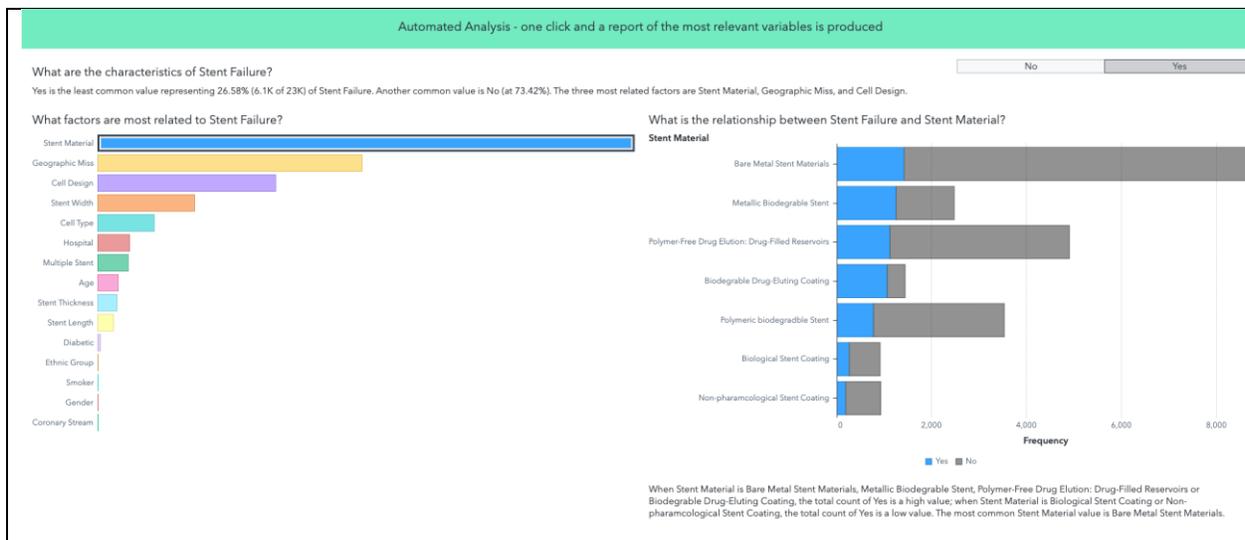
DESCRIBE THE ANALYSIS

SAS Visual Analytics has several analytical and statistical methods to help identify root causes (Display 6). If you are not sure where to start, the best approach is to use the automated explanation.



Display 6. Analytics & Statistical Methods

SAS Viya offers an important feature to facilitate data exploration: Automated Explanation (AE) uses machine learning to suggest business intelligence visualizations for a measure or category outcome variable, producing interactive results. For the chosen response variable, AE generates a summary of the outcomes, relative importance scores for each underlying factor, and a plot reflecting the relationship between a selected underlying factor and the outcome variable (Table 1). AE also reports the specific combinations, denoted as groups, of influential data items most and least likely to yield a given response.



Display 7. Automated Explanations

Outcome Variable Type	Underlying Factor: Measure	Underlying Factor: Category
Measure Data	Scatter Plot or Heat Map	Bar Chart
Category Data	Histogram	Stacked Bar Chart

Table 1. Automated Explanation Relationship Plots

SAS Visual Analytics screens the data items to specify underlying factors according to certain selection criteria (Table 2). For each variable, AE decides on an optimal one-level decision tree splitting, then ranks the variables based on the decision trees using a relative importance score. Lastly, high and low groups are determined by running several additional decision trees on the outcome variable; each group represents a specific leaf from one of those decision trees. After creating an automated explanation, we can take the exploration further by doing the following:

- Change or duplicate the AE to another object (and vice versa)
- Derive a new data item or create a new object based on one or more selected groups

Excluded From Consideration as Underlying Factors	Added as Factors but Later Removed
Aggregated Measures	Duplicated Data Items
Automatically Computed Values (Frequency, Hierarchies, Filter Items, Interaction Effects, Spline Effects)	ID Data Items
Date & Time Values	Items Identical to Outcome
	Items w/ Many Missing Values

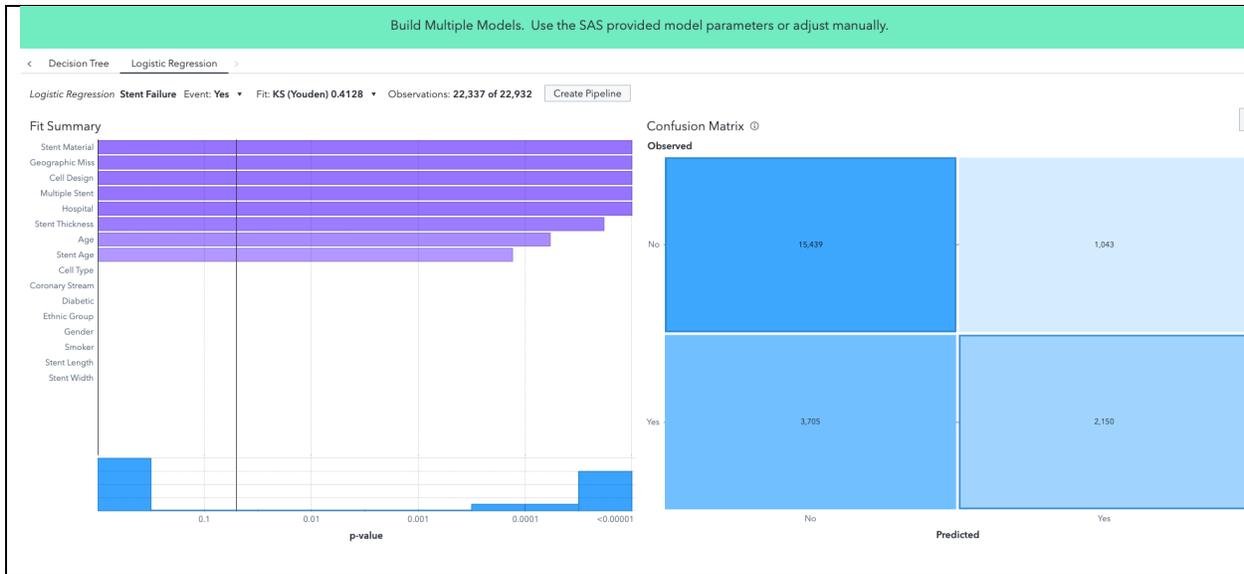
Table 2. Automated Explanation Underlying Factor Selection Criteria

Using Automated Explanation (as well as Automated Prediction) saves time and increases productivity. AE explains basic data-driven insights, allowing data scientists to focus on solving more complicated problems and creating higher-value models. Moreover, because AE accommodates nontechnical users, decision makers become increasingly insight driven, which aids in developing effective policy.

Multiple best practices to consider regarding automated explanation and prediction in SAS Viya:

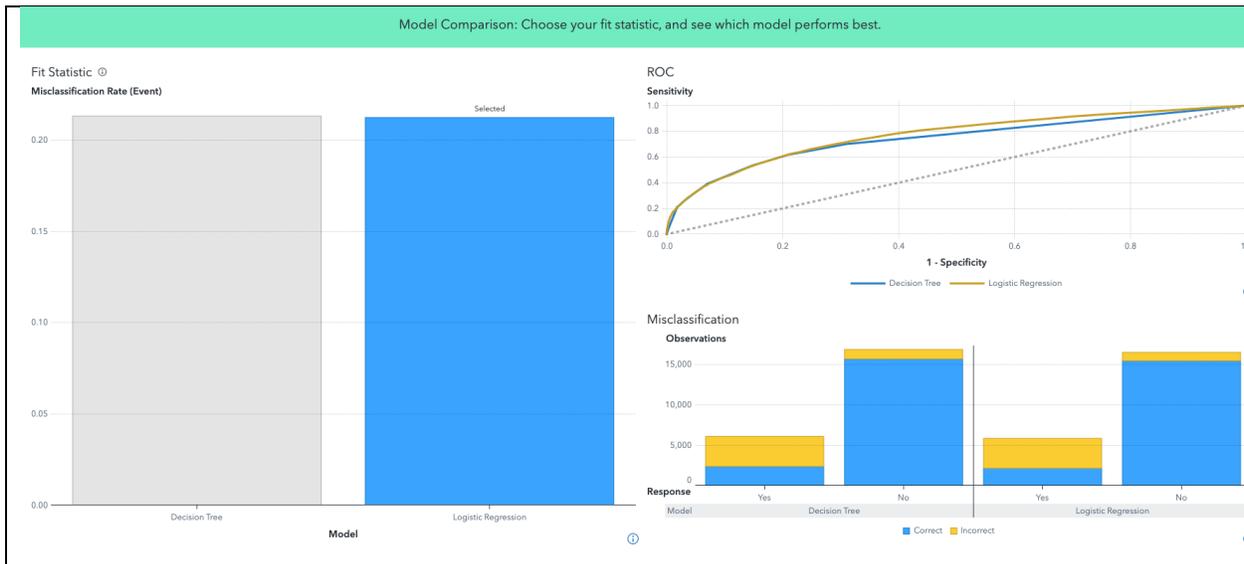
- Ensure all variables are imported in the correct format
- Ensure there are no duplicated underlying factors (such as two copies of the same variable, one formatted as a measure and the other as a category)

For users who are more comfortable with using statistical models, Visual Statistics offers several choices that can be produced with no coding required. In fact, you can duplicate the automated analysis as a logistic regression with a single click that will keep the same response and explanatory variables (Display 8).



Display 8. Logistic regression output.

The options section of the logistic regression allows you to change the hyperparameters for the regression (e.g., variable selection methods). If you make multiple models, you can compare the results based on a wide selection of possible model selection statistics. (Display 9)



Display 9. Model comparison.

SUMMARIZE FINDINGS

To wrap up the presentation, remind the audience of the problem and the key findings. These can have more information than the key findings earlier in the presentation but take care to not overwhelm here. A summary of next steps for how to move forward is also appropriate.

CONCLUSION

SAS Visual Statistics is a powerful tool to help you tell data stories and answer business questions. It has the ability to combine text, images, data visualizations and analytic results in one package that is easy to consume by a non-technical audience.

RECOMMENDED READING

1. SAS Institute Inc. "SAS Visual Analytics for SAS Viya: Chapter 5". Available [Chapter 5: Advanced Topics – Automated Explanations | SAS Visual Analytics for SAS Viya \(oreilly.com\)](#). Last Modified December 2020.
2. Styll, R. "AI is Coming for Your BI: Automated Analysis in SAS® Visual Analytics". Available <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3526-2019.pdf>. Last Modified March 28, 2019.
3. SAS Software. "Automated Analysis with SAS Visual Analytics". Available <https://www.youtube.com/watch?v=Jl-zZ7LhXfE>. Last Modified December 19, 2018.
4. SAS Institute Inc. "Working with Automated Explanation Objects". Available <https://go.documentation.sas.com/doc/en/vacdc/8.4/vaobj/n0ewwfd6udhv7qn1nropd18lof9j.htm>. Last Modified August 19, 2019.
5. Carey, M. "Take customer care to the next level with automated prediction in SAS Visual Analytics". Available <https://blogs.sas.com/content/sqf/2020/08/17/automated-prediction-customer-care/>. Last Modified August 17, 2020.
6. SAS Institute Inc. "Working with Automated Prediction Objects". Available https://go.documentation.sas.com/doc/fr/vacdc/v_002/vaobj/p0sytm7v7ft0c6n1x2a7k5i8sa9b.htm. Last Modified January 25, 2021.

CONTACT INFORMATION

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

Jim Box

SAS Institute

Jim.box@sas.com

Laura Watson

SAS Institute

Laura.watson@sas.com

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

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