sas innovate

Easily Turn Your Automated Explanation into a Predictive Model

Danny Modlin

#SASinnovate

Abstract

The Automated Explanation in SAS® Visual Analytics is a powerful and easy way to start your journey into building predictive models. We'll start by creating an Automated Explanation to analyze our data and pick up on key drivers that appear. We will then easily move into building a predictive model using SAS® Visual Statistics. Data Scientists and Business Analyst will see how easy it is to take that next leap in their data analysis. Code will not be harmed, examined, nor executed during this workshop.

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

Lesson 1 Easily Turn Your Automated Explanation into a Predictive Model

1.1 Introduction 1-5

1.2 Hands-on Workshop 1-10

Demonstration: Build an Automated Explanation and an Automated Prediction 1-11

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

1.1 Introduction

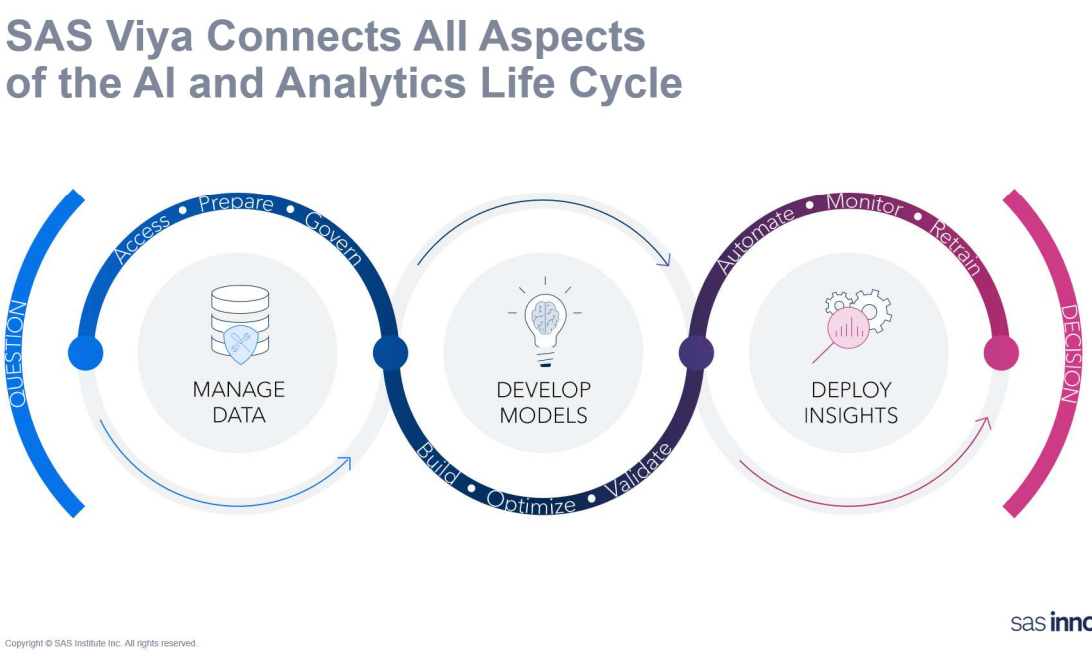
Objectives

- Introduce SAS Viya.
- Introduce SAS Visual Analytics in SAS Viya.
- Discuss automated explanation basics.
- Demonstrate an automated explanation and predictive modeling.

Copyright © SAS Institute Inc. All rights reserved.

sas innovate 2025

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.



SAS Viya is a cloud-enabled, in-memory analytics engine that provides quick, accurate, and reliable analytical insights. In SAS Viya, the SAS High-Performance Architecture enables the high-performance analytics engine. The CAS In-Memory Engine continues the ability to perform processing in memory and the ability to distribute processing across nodes in a cluster. The CAS In-Memory Engine adds highly efficient node-to-node communication and uses an algorithm to determine the optimal number of nodes for a given job.

SAS Cloud Analytic Services, or CAS, is a server that provides the cloud-based run-time environment for data management and analytics with SAS. By *run-time environment*, we refer to the combination of hardware and software where data management and analytics take place.

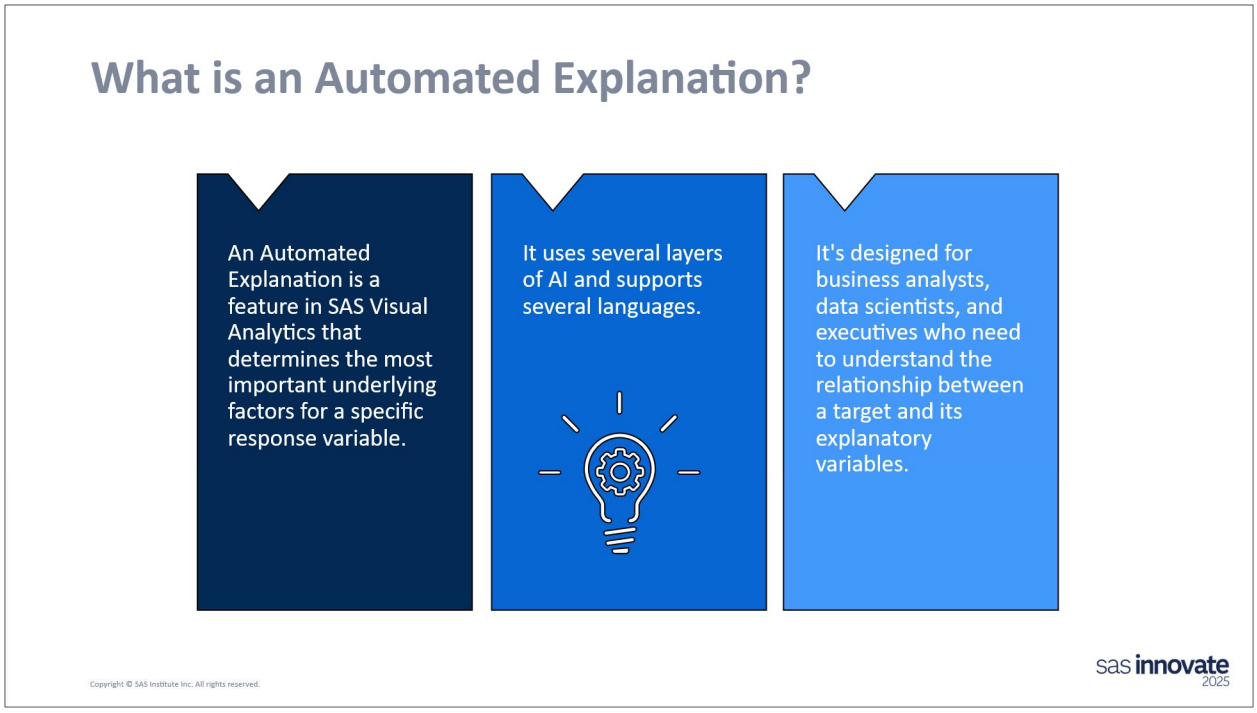
The server can run on a single machine or as a distributed server on multiple machines. The distributed server consists of one controller and one or more workers. This architecture is often referred to as a *massively parallel processing architecture*. For both modes, the server is multi-threaded for high-performance analytics.

The distributed server has a communication layer that supports fault tolerance. A distributed server can continue processing requests even after losing connectivity to some nodes. The communication layer also enables you to remove or add nodes from a server while it is running.

One of the design principles of the server is to handle large problems and to work with tables that exceed the memory capacity of the environment. In order to address this principle, data in the server is managed in blocks. Whenever needed, the server caches the blocks on disk. It is this feature that enables the server to manage memory efficiently, handle large data volumes, and remain responsive to requests.

You can use a variety of interfaces to interact with the CAS In-Memory Engine. These interfaces include SAS Studio, which is a browser-based interface for writing SAS code. You can also use programming interfaces for R, Python, Java, and Lua to access this CAS functionality. In addition, you can continue to submit SAS code in batch mode.

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.



Visual Analytics uses SAS High-Performance technologies to accelerate analytic computations, which helps you derive value from massive amounts of data. This gives you the power to solve difficult problems, improve business performance, predict future performance, and mitigate risk rapidly and confidently. You can import data of any size into Visual Analytics to quickly identify trends and patterns that you might not have noticed before.

Then, you can take your analysis a step further by creating powerful statistical models on the patterns that you discovered (with SAS Visual Statistics). After you identify those patterns and get a better feel for your data, you can create reports or dashboards to share with anyone, anywhere via the web or a mobile device.

SAS Visual Analytics helps you to visualize and discover relevant relationships in your data. You can create and share interactive reports and dashboards and use self-service analytics to quickly assess probable outcomes for smarter, more data-driven decisions.

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

Who would use and Automated Explanation?

An Automated Explanation is useful for anyone who needs to understand the relationship between a target variable and its explanatory variables.

This includes business analysts, data scientists, and high-level executives.

For example, the head of Customer Loyalty at a company might use it to understand which factors affect customer satisfaction.

Copyright © SAS Institute Inc. All rights reserved.

sas innovate 2025

What can you do with an Automated Explanation?

An Automated Explanation quickly builds a series of easily interpretable visualizations along with automatically generated storylines.

It reveals the most important underlying features for a target variable.

It provides a relative importance score for each underlying factor. The most important underlying factor is assigned a score of 1, and all other scores are proportional to that value.

Copyright © SAS Institute Inc. All rights reserved.

sas innovate 2025

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

What else can you do with an Automated Explanation?

An Automated Explanation can be used to explore high and low groups and examine the relationship between the response and the underlying factor.

It can be used to predict responses based on adjusted values of underlying factors.

It is a good starting point for building more complex models. It is easy to convert to many other predictive models available.

Copyright © SAS Institute Inc. All rights reserved.

sas innovate 2025

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

1.2 Hands-On Workshop

In this workshop, you use the data set **vs_bank**. This data set consists of observations taken from account holders at a large financial services firm. The accounts represent consumers of home equity lines of credit, automobile loans, and other short- to medium-term credit instruments. Appropriate data cleansing has already been applied, so we can begin with statistical modeling. The target variables relate to whether that account holder purchased a new product from the bank in the past year. The data sets contain more than 1 million rows and 24 columns. A list of variables and their labels is shown below.

Target Variable

B_TGT	New Product (Binary)

Categorical Inputs

CAT_INPUT1	Account Activity Level
CAT_INPUT2	Customer Value Level

Interval Inputs

RFM1	Average Sales Past Three Years
RFM2	Average Sales Lifetime
RFM3	Avg Sales Past Three Years Dir Promo Resp
RFM4	Last Product Purchase Amount
RFM5	Count Purchased Past Three Years
RFM6	Count Purchased Lifetime
RFM7	Count Prchsd Past Three Years Dir Promo Resp
RFM8	Count Prchsd Lifetime Dir Promo Resp
RFM9	Months Since Last Purchase
RFM10	Count Total Promos Past Year
RFM11	Count Direct Promos Past Year
RFM12	Customer Tenure

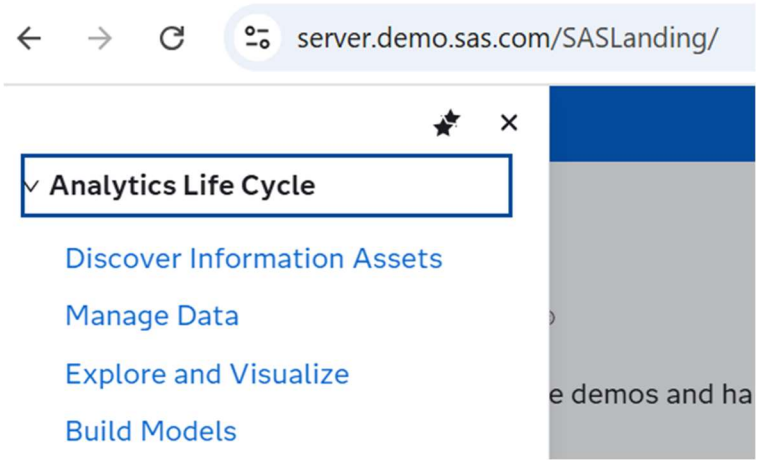
Note: Other variables, not listed here, are also included in the data set. Variables with the prefix **I_** are imputed. Variables with the prefix **RI_** are imputed and replaced. Variables with the prefix **LOGI_** are imputed and log transformed. Variables with the prefix **DEMOG_** are demographic inputs.

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

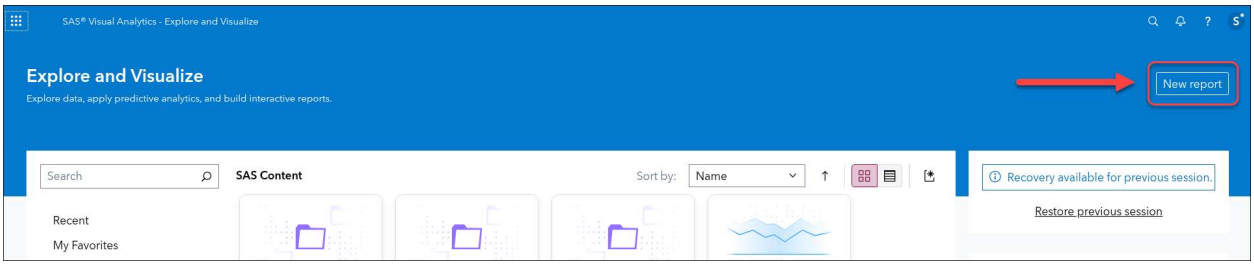
Building a Logistic Regression Model

This demonstration illustrates how to build a logistic regression model in SAS Visual Analytics. The demonstration uses the **vs_bank** data to model whether a customer contracted for at least one product in the previous campaign season. You create a binary logistic regression with both categorical and continuous explanatory variables. You then perform model validation and variable selection.

1. From the desktop, open Google Chrome.
2. From Google Chrome toolbar, select **SAS Landing**.
3. Enter **student** in the **User ID** field, if necessary.
4. Enter **Metadata0** in the **Password** field, if necessary.
5. Select **Sign in**.
6. If requested to save the password, select **Save**.
7. Select **YES** when asked about assumable groups.
8. Access the applications menu in the upper left of the window and select **Explore and Visualize**.

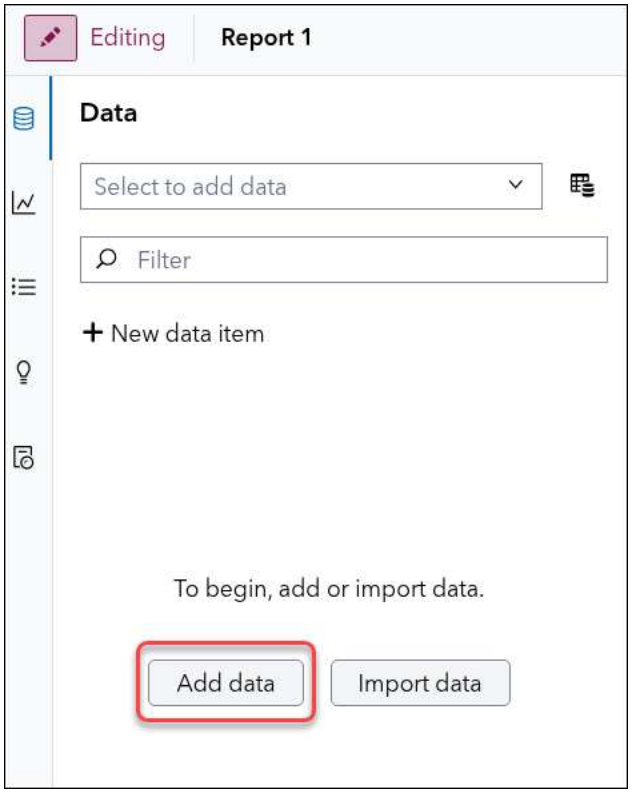


9. Load the SAS data set into CAS.
   a. In the Explore and Visualize window, click **New report**.



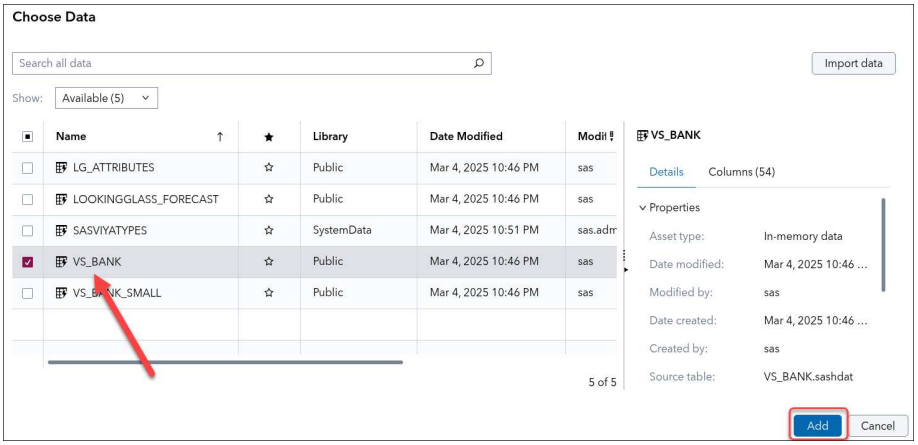
Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

b. On the Data tab on the left side of the screen, select **Add data** to load an in-memory table to SAS Visual Analytics.



Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

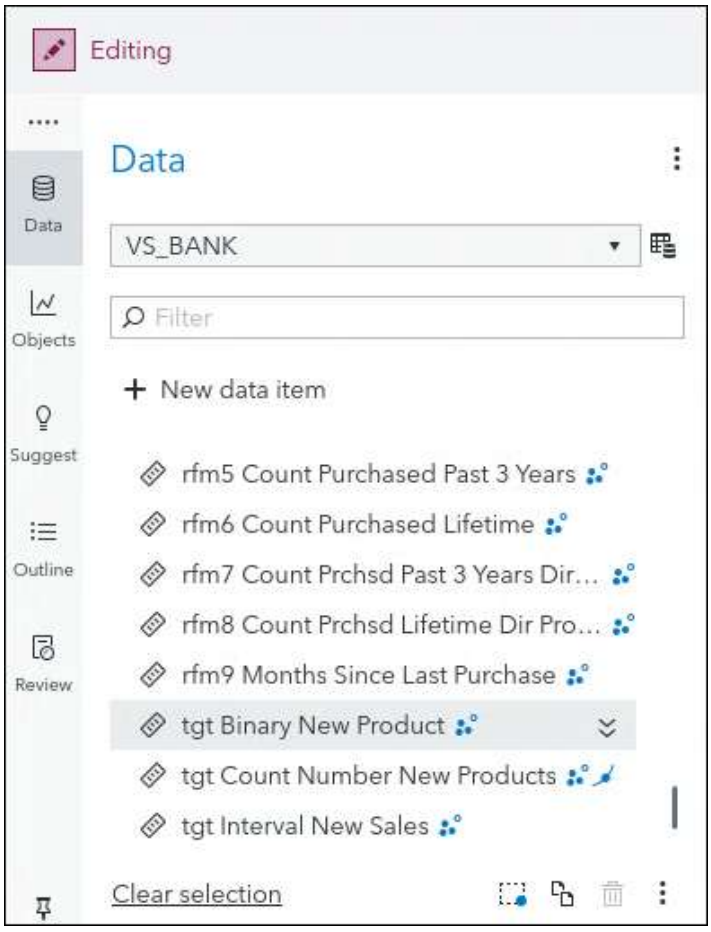
10. In the Choose Data window, select **VS_Bank > Add**.



The table is now available to the newly created report.

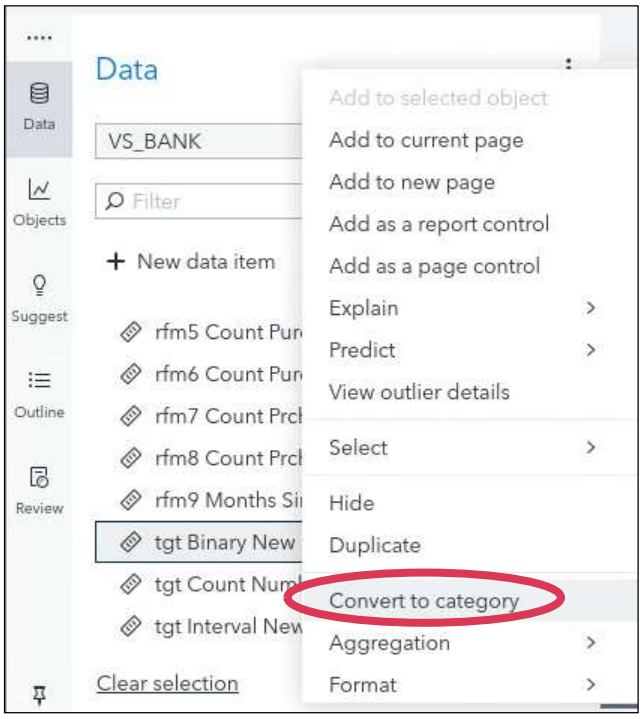
11. On the Data tab on the left of the screen, scroll down and find the measure **tgt Binary New Product**.

The variable **tgt Binary New Product** (**b_tgt**) is the primary dependent variable for categorical response modeling in this workshop. It is a binary flag that codes responders with 1 and non-responders with 0. Because it is numeric, it is treated as interval valued by default.

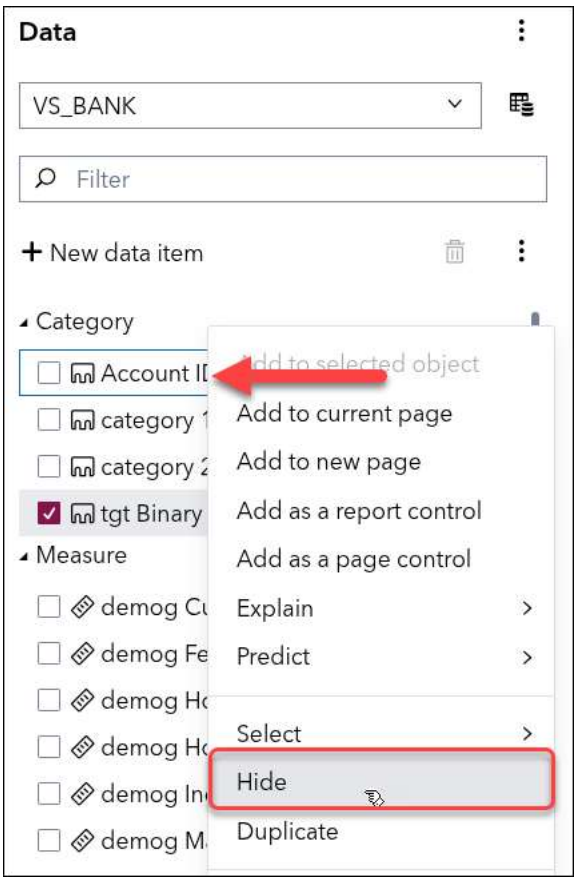


Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

12. Right-click **tgt Binary New Product** and select **Convert to category**.

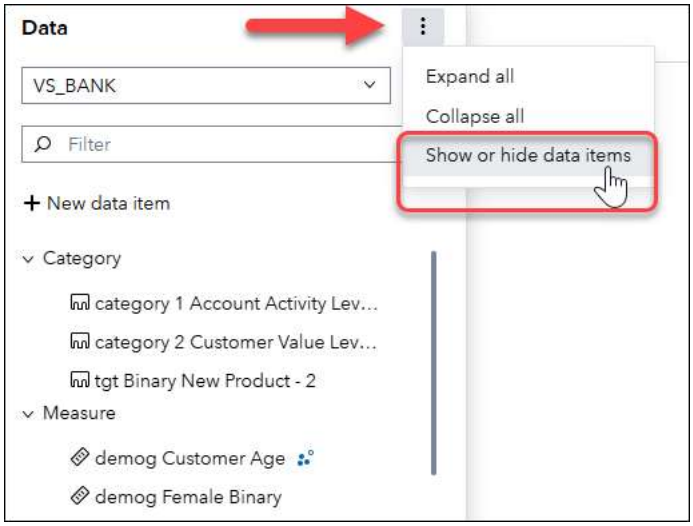


13. We need to hide several inputs so that they will not be included in our automated explanation. We will start by hiding the account id. Right-click **Account ID** and select **Hide**.



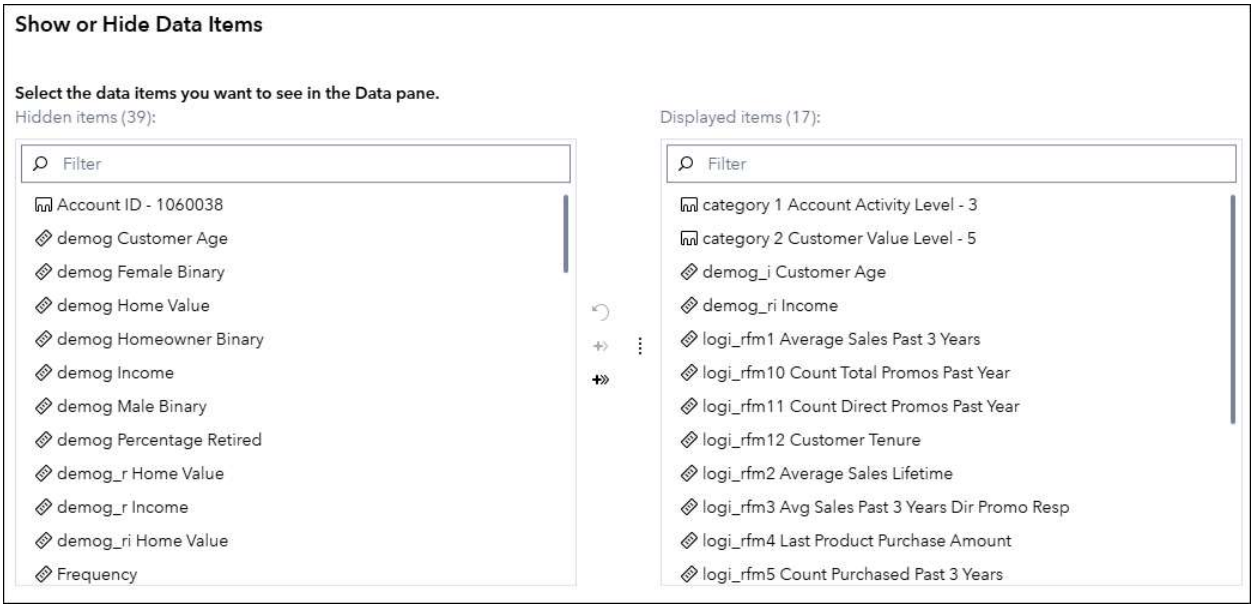
Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

14. Next we will hide several variables at one time. Select **Data menu > Show or hide data items**.



15. In the Show or Hide Data Items window, move all the measures over to the hidden items column except for:
    a. demog_i Customer Age
    b. demog_ri income
    c. all logi_ variables

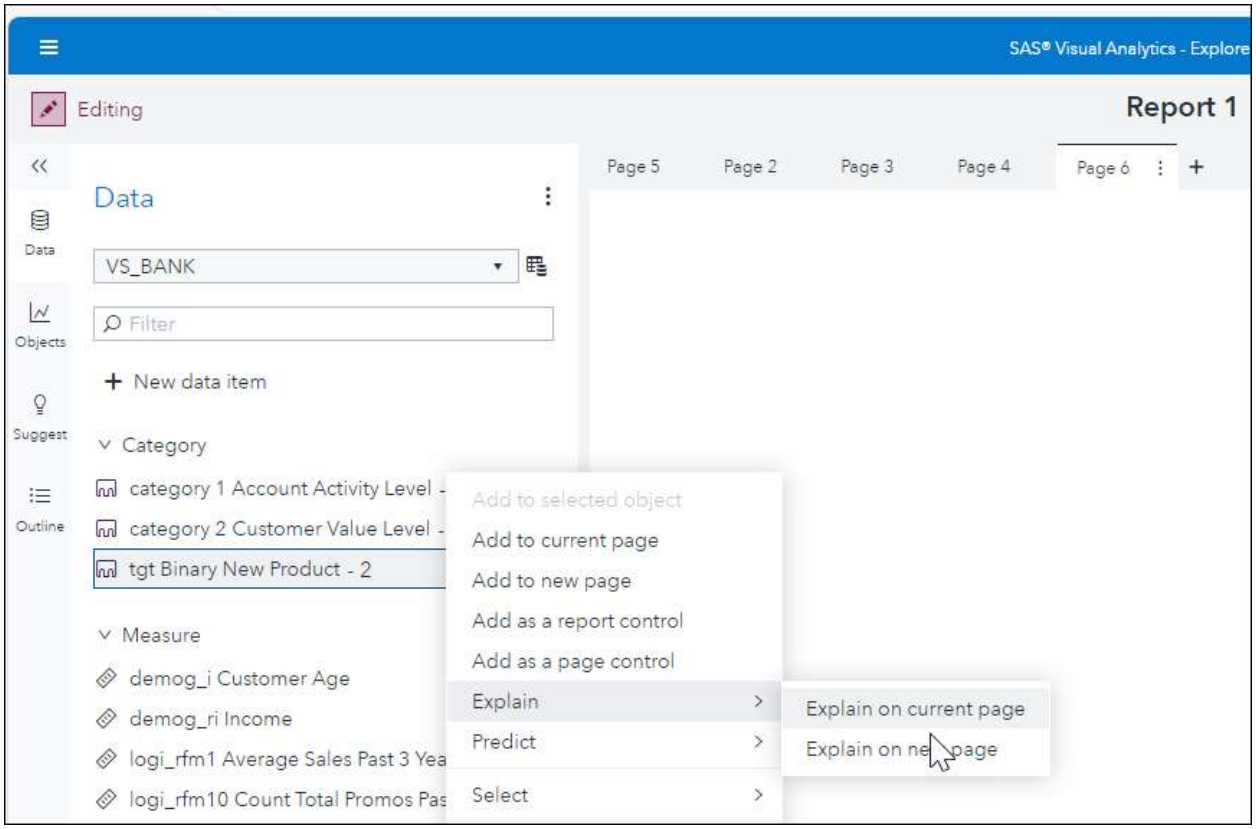
    Important Note: Do **not** move the 2 remaining categorical variables to be hidden. Do **not** move the target variable to be hidden.



16. You should have 17 Displayed items. Click **OK** to close the Show or Hide Data Items window.
17. The data set contains more than one million rows and (is now filtered to) 16 explanatory columns plus one target variable. We will see more detail on the variables in our exploration, but they contain demographic information, account activity level, customer value level and various purchase behaviors.

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

A quick and easy way to begin exploring my data would be to create an automated explanation. The automated explanation reveals the most important underlying features for a target variable. In this example, I'm trying to understand whether an account will make a purchase (or not). Let's easily create that explanation in a report with one click. I right-click on the target variable **tgt Binary New Product** and select **Explain > Explain on current page**.

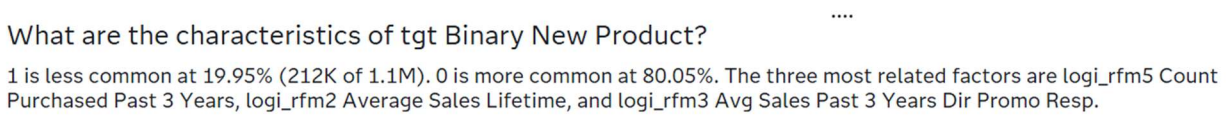


18. The resulting report reveals that my target variable has an 80% chance of being a 0. In other words, many of my customers were non-purchasers.

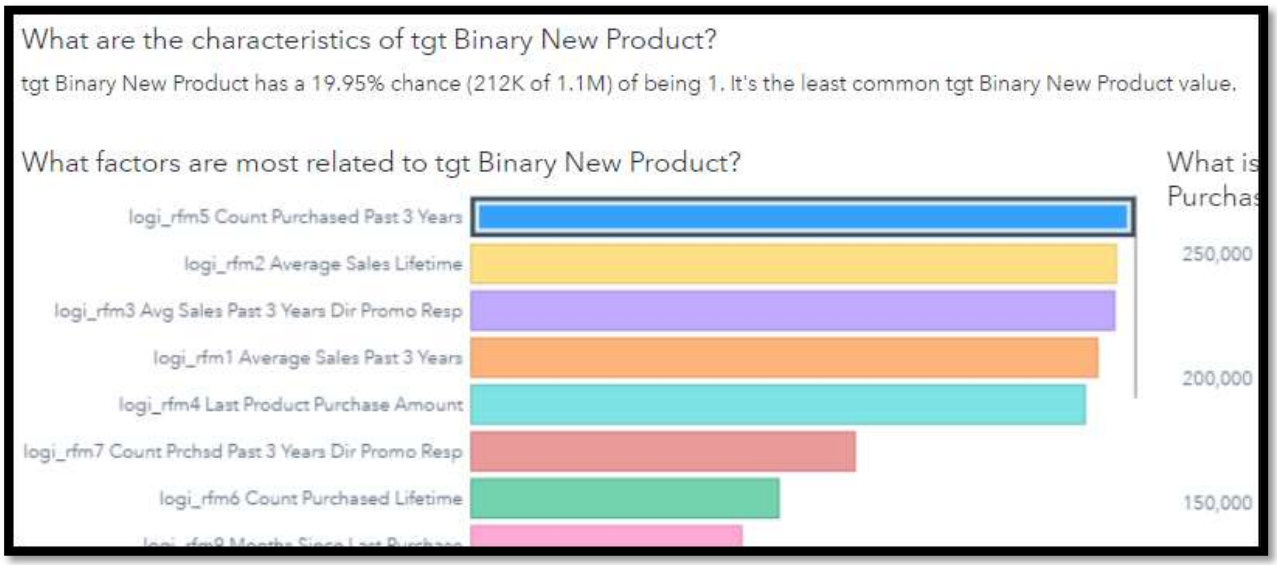
    Much of the report aims at explaining the most common value of 0 (non-purchasers), but honestly, we are more interested in the behavior of the purchasers (value of 1). Let's update the chart by selecting **1** in the button bar.

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

19. From the resulting report, I begin my data exploration and discover all kinds of interesting information about the target variable of customer purchase. You'll notice that the summary bar along the top has been updated to show that approximately 20% of the customers made a purchase (value of 1).

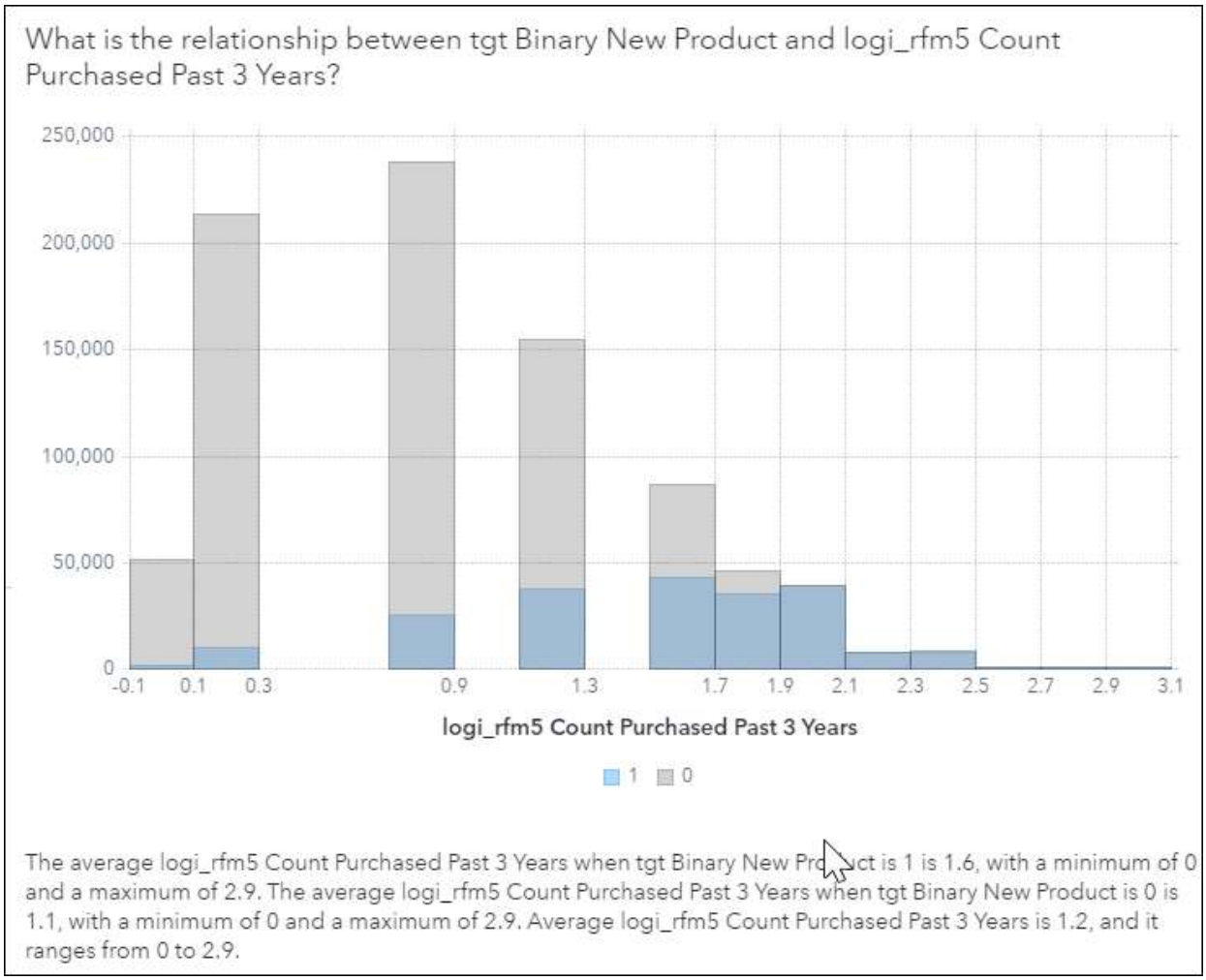


Then we can see under "What factors are most related to tgt Binary New Product?" that the following three variables are the most related factors: count purchased over the past 3 years, average sales over the lifetime, average sales over the past 3 years in response to a direct promotion. Of course, it makes sense that these three factors could have a large effect upon whether a customer would make a purchase or not. Notice that the top bar is already selected.



Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

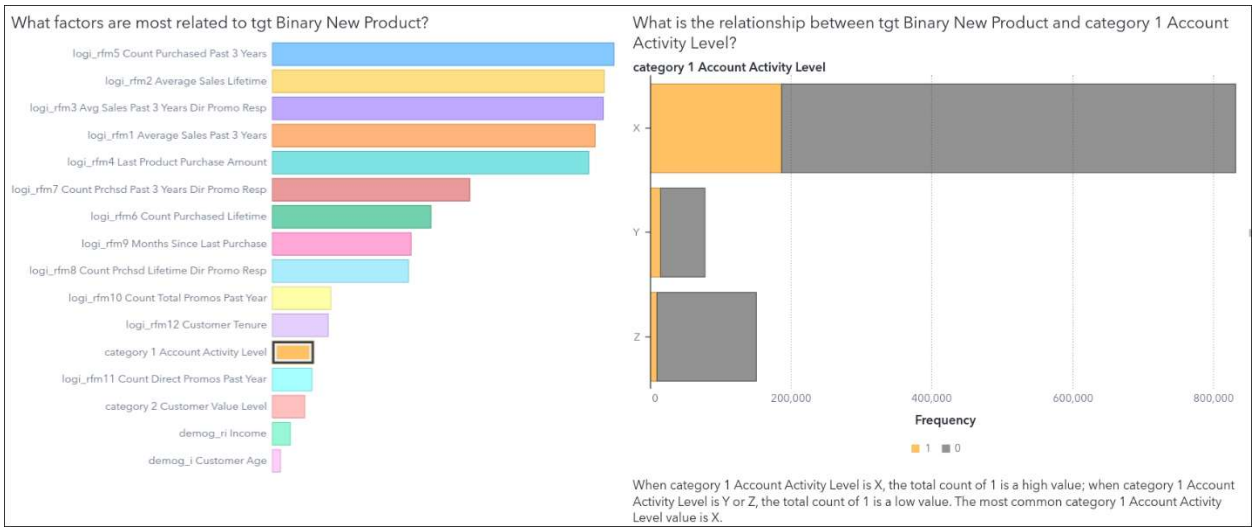
20. It would be interesting to understand the relationship between this top factor (count purchased over the past 3 years) and our binary target variable. Fortunately, we already have an automatically generated chart to help us. Let's examine "What is the relationship between tgt Binary New Product and logi_rfm5 Count Purchased Past 3 years?"



We can see that for purchasers; the average number of products bought over the past 3 years is about 1.6. Keep in mind that our data was transformed, so customers bought approximately 5 products on average.

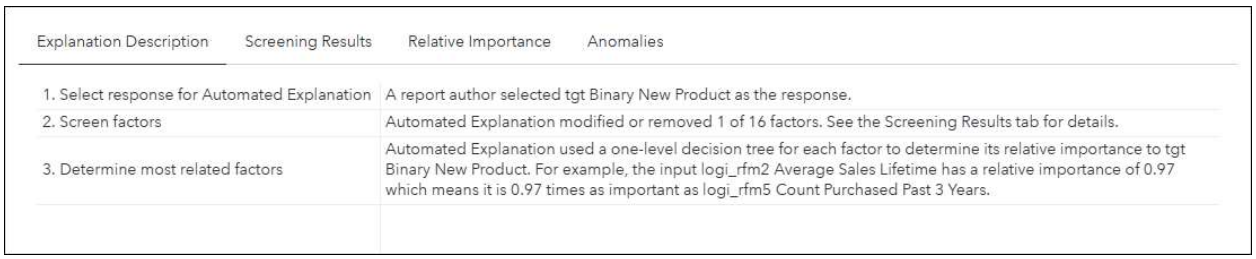
Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

21. What happens to the automatic explanation if we are interested in a categorical input? To investigate the relationship between the account activity level and our target, select the **category 1 Account Activity Level** bar. The relationship chart automatically updates. It makes sense that the accounts with the highest activity (value of “x”) have most of our purchasers.



22. Before continuing with examining the rest of the automated explanation, let’s reset the chart by re-selecting the top bar of the most related factors chart: **logi_rfm5 Count Purchased Past 3 Years**.

23. On the report, click (**Maximize**) to see the Details table.



The Explanation Description includes: the selected response for the automated explanation, the screen factors, and how the most related factors were determined. In this case, a one-level decision tree for each factor was used to determine relative importance.

The Screening Results show you steps that the Automated Explanation took to prepare your data for its analysis. It might remove some variables (for example, those that have a high number of missing values). It converts numeric columns with few distinct values to categories. It might even remove geographic variables like latitudes and longitudes from the analysis.

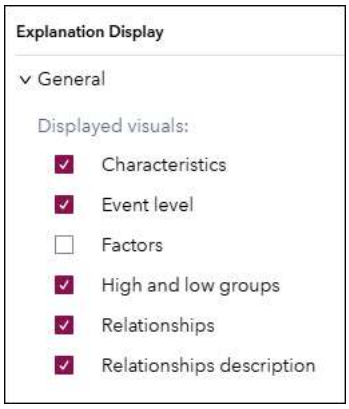
Relative Importance shows you the values used in the most related factors chart.

Anomalies reveal any anomalies detected during the automated explanation.

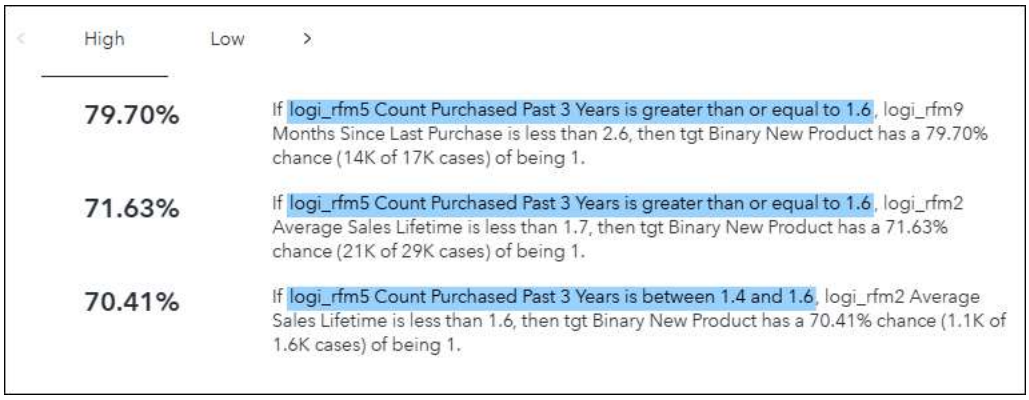
24. In the upper right corner of the report, click (**Restore**) from the object toolbar to close the Details table and exit maximize mode.

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

25. Before we complete our investigation of the automated explanation, let's turn on an option to show us the most likely and least likely groups of purchasers. In the Options pane, turn on **High and low groups**. (Turn off **Factors** for more real estate.)

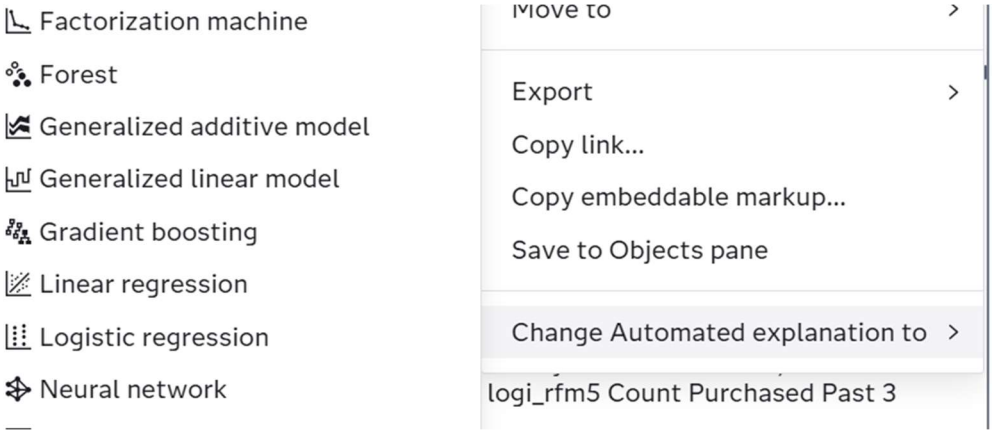


26. A new visual appears on the canvas of our automated explanation. On the High tab we are presented with the top three groups that are most likely to make a purchase. Let's examine the first group which has an almost 80% predicted probability of making a purchase. If the count purchased over the past 3 years is greater than or equal to 1.6 **and** it has been less than 2.6 months since the last purchase, then a customer is very likely (79.70%) to make a purchase. In case you are curious, there are more decision trees being created in the background to give us all this wonderful information.



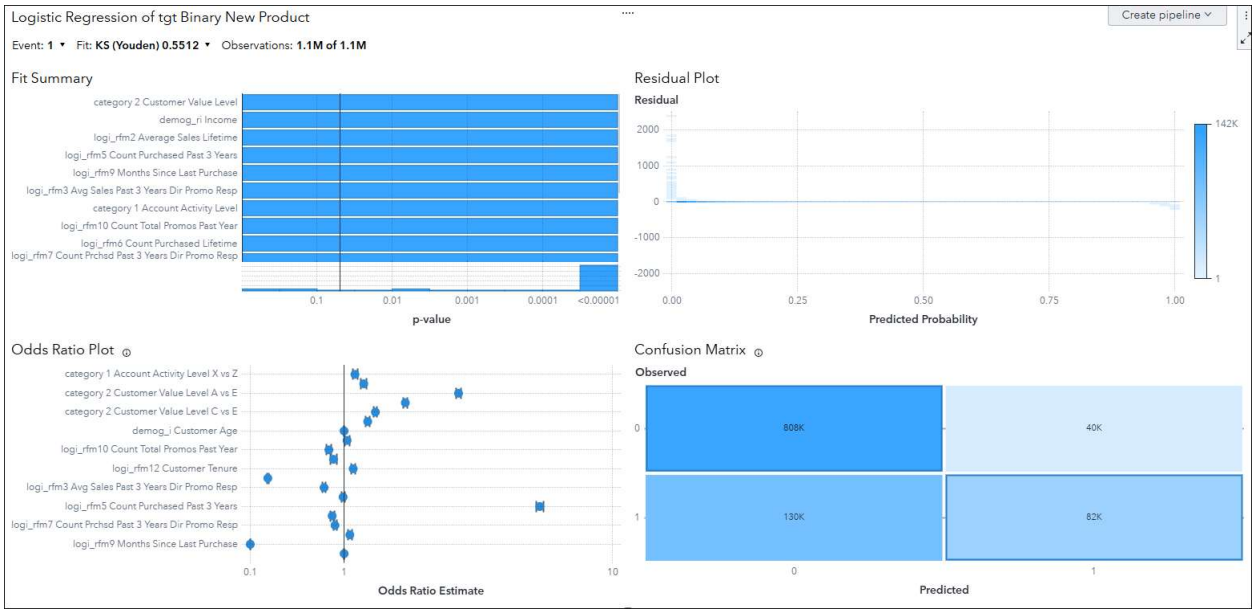
Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

27. Now that we've done a bit of data exploration, let's quickly and easily see how we can build a predictive model. I open the **Object menu** and select **Change Automated explanation to > Logistic Regression**.



Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

28. I am still astounded at how easy it is to create models in SAS Viya. As you've seen, just a couple of clicks and I already built a predictive model! As a data scientist, being able to quickly explore data and efficiently build models are important skills. Let's take a quick look at the logistic regression results.



From the Summary Bar at the top, we can see that the KS statistic is 0.5512. We could use that model fit statistic to compare it to another model and determine which model had the higher value. The Fit Summary panel reveals that 14 of the 16 effects are significant at a .05 significance level. The two predictors at the bottom of the chart are not significant at .05. There is a large odds ratio of 6 for *Count purchased over the last three years*. This means that for each additional product a customer purchases, the customer is 6 times more likely to be a purchaser. The Residual Plot shows that there do not appear to be any outliers in the residual data. And finally, the Confusion Matrix reveals that our model correctly identified 81,853 of the purchasers and 808,181 of the non-purchasers.

End of Demonstration

Copyright © 2026, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.