

AI is Coming for Your BI: Automated Analysis in SAS® Visual Analytics

Rick Styll, SAS Institute Inc., Cary, NC

ABSTRACT

Automated analysis, a new feature in SAS® Visual Analytics, uses machine learning to automatically suggest business intelligence (BI) visualizations for you. Powered by the SAS® Platform, this analysis is performed quickly so that results are interactive. Once you choose a variable you are interested in, the most relevant factors are automatically identified and easy to compare. As a business analyst or manager, you can immediately understand the results in a report by using the automated narratives and familiar BI visualizations.

INTRODUCTION

When a key metric is higher or lower than expected, someone is asked to explain why. Typical dashboards and reports show you what happened, what is higher or lower than expected, but not why. Today, there is an abundance of data. However, even with all the attention lately on training data scientists, the people with requisite skills and the time required to find what is driving these unexpected outcomes are both scarce.

This growth in data will only continue. It's widely recognized that making decisions and taking actions based on data-driven insights will lead to better outcomes. Can we train more people to become data scientists? Yes, but there isn't a way to train enough people fast enough. Data is growing faster.

Just as important, there isn't enough time or specialists to explain all of the predictive modeling results to the non-technical business professionals that can make or influence changes. You might be quite technical, but most business people in your organization depend on IT and technical business professionals, like you, to not only create but to explain data-driven insights. I believe that automating predictive modeling as well as the presentation and explanation of the modeling results is the best way to assist the vast majority of people in any organization to understand why their metrics are too high or too low. With automated analysis, more people can identify the right changes to make in order to improve outcomes faster.

Automated analysis takes a fraction of the time with less data science skill leading to the following benefits:

- Managers (and those supporting them directly) will become more independent and insight-driven.
- Business Analysts will use automation to be more productive.
- Expert data scientists will have more time to create high-value models to solve more complex problems.

In September 2018, SAS delivered a new feature in SAS Visual Analytics 8.3 called automated analysis. In this paper, we'll take a close look at this new feature and how it can help you and many others in your organization quickly understand what's driving your data.

WHY AUTOMATION OF ANALYSIS NOW?

Data and predictive analytics are not new. So what makes now the right time to consider automation of predictive analytics to assist people who are making or influencing changes to improve outcomes? There is a confluence of trends, including the following, that is driving the demand, feasibility, and availability of automated analysis:

- A tidal wave of new data (IoT sensors, data from apps, external data sources, unstructured data, and so on) is upon us.
- The awareness of the value predictive analytics enables has soared.
- Data scientists are in limited supply.
- High-performance computing now enables interactive modeling on very large volumes of data.
- Cloud infrastructure and technology has reduced costs and deployment times dramatically.
- Natural language processing is making conversational analytics a reality.

For most of us, we've witnessed innovations that have automated many manual tasks at various points in our lifetimes. ATMs (Automated Teller Machines), email, the internet, smartphones, and social media are just some of the more salient ones. Some more recent innovations include using drones to deliver packages, self-driving vehicles, and 3-D printers, but these have not been widely adopted, yet.

In the book, *Abundance: The Future Is Better Than You Think*, authors Peter Diamandis and Steven Kotler share examples of technologies that have gone from being narrowly adopted or scarce initially to ones that are abundant. Technologies start off being scarce because they are too expensive, not accessible, and often too difficult-to-use for anyone but specialists. Technologies become abundant later, sometimes much later, once the technology is affordable, accessible, and easy-to-use for most everyone (Diamandis and Kotler, 2012). Predictive analytics has gone through this cycle over several years. However, the adoption of predictive analytics is approaching a tipping point. Only recently has the cost, accessibility, and usability (based on some of the factors in the bulleted list above) made it possible for predictive analytics to be adopted by the masses. Automation of the manual steps to data preparation, predictive modeling, visualization, and explanation will make predictive analytics massively adopted in the not-so-distant future.

Automation can be seen as a threat to people's skills and perhaps even their jobs. According to legend, the African American folk hero and steel driver John Henry raced against a steam-powered rock drilling machine when building a railroad tunnel. He won the race, but he died from stress shortly after he won with his hammer in his hand. Automated analysis is not being driven by a need to reduce the number of data scientists or even stress them out. It is being driven by a need to help them and others with the abundance of data and the limited supply of data scientists' skills and time. Automated analysis can help managers and analysts, but it cannot replace these roles nor the role of a data scientist. ATMs did not eliminate bank tellers, loan officers, and branch managers. It freed them up from having to handle all transactions manually so that they can be more productive, focusing on higher value services, creating better customer experiences, and cross selling other bank offerings. Although automation often reduces or even eliminates tedious manual tasks that people once had to perform, it enables outcomes not achievable before, and allows specialists to focus on more rewarding aspects of their work.

AUTOMATED ANALYSIS IN SAS VISUAL ANALYTICS

Starting with the September 2018 update of SAS Visual Analytics 8.3, you can get a jump start on your data discovery with automated analysis. No additional licenses are required. First, you simply select a data item representing an outcome that you want to understand better and SAS Visual Analytics automatically analyzes all of the columns and rows in the underlying data source to determine which data items are driving the outcome and how. You can quickly see relationships in any data that you have access to. If your starting point is an existing report or dashboard, you can use automated analysis in SAS Visual Analytics to understand why certain results have occurred and what is likely to happen next. Automated analysis works whether you need to learn about an outcome that is a numeric measure or a descriptive category, such as an “unsubscribed” customer status. Either way, the results from the automated analysis are presented with simple, natural language explanations and easy-to-understand visualizations so that you, or almost anyone, can quickly grasp the essential insights.

Imagine for a moment that you work for a video game publisher and you have just created the following high-level report in SAS Visual Analytics. You can see revenue by video game genre in the donut chart on your left. You can see some details about the genre and class combinations in the list table to your right, ordered by revenue highest to lowest. You decide to show your new report to your manager and colleagues during a weekly team meeting. After a brief team discussion your manager asks, “What can we change to increase overall game revenue?”

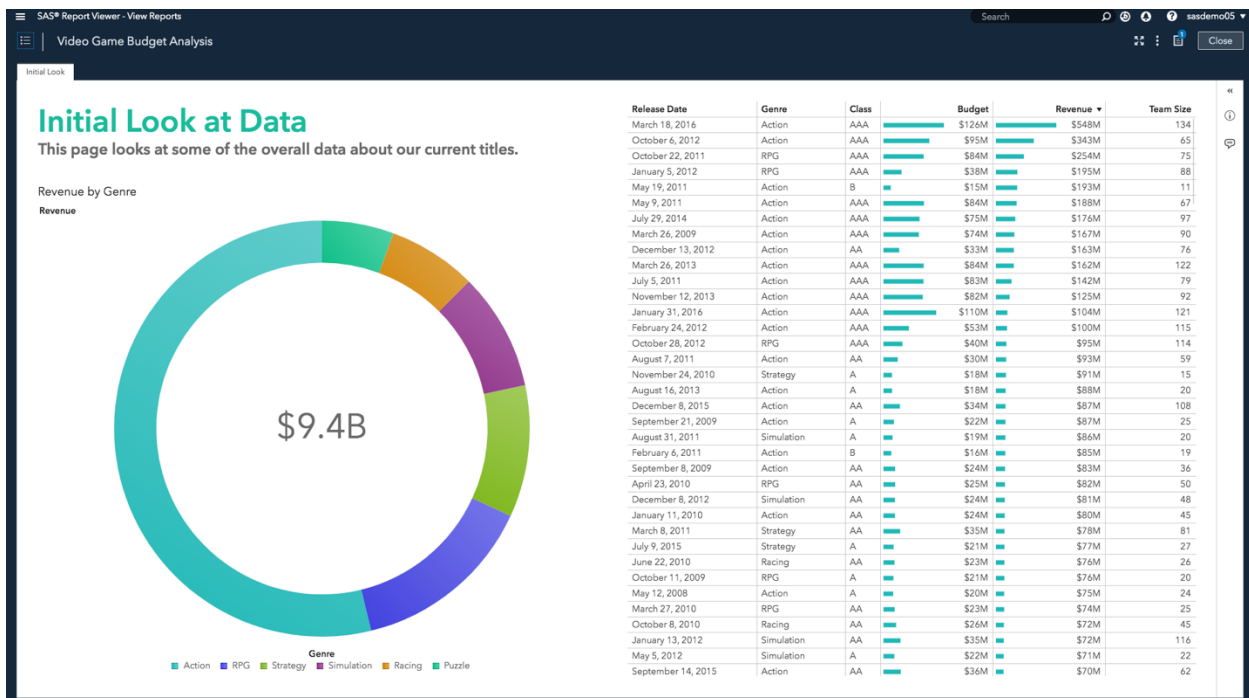


Figure 1. An Initial View of a Fictitious Video Game Publisher’s Data

You weren't prepared for this question quite yet. To better understand what is driving revenue, you could manually add a decision tree to a page, perhaps a forecast, and many other visualization types in SAS Visual Analytics, but there are so many columns in your data to choose from. To accomplish this, you could reply to your manager's question that you need the rest of the day to build the analysis and type an explanation in an email to describe the results. Like most of your teammates, your manager isn't proficient interpreting decision trees. Instead, you decide to use automated analysis in SAS Visual Analytics on-the-fly in front of your entire team to quickly show and explain which factors are most important in predicting revenue for video games and the values for those factors that are most important.

AUTOMATED ANALYSIS FOR A MEASURE

Automated analysis quickly determines the most important underlying factors for a specific outcome. Often, the outcome of interest is a key metric (for example, hospital patient readmission rate, units of inventory, percentage of sales quota, and so on). In this video game publisher example, we'll select the **Revenue** data item, as shown in Figure 2. When you right-click **Revenue**, you will see **Analyze** in the menu with a choice to display the analysis results on the current page or a new page.

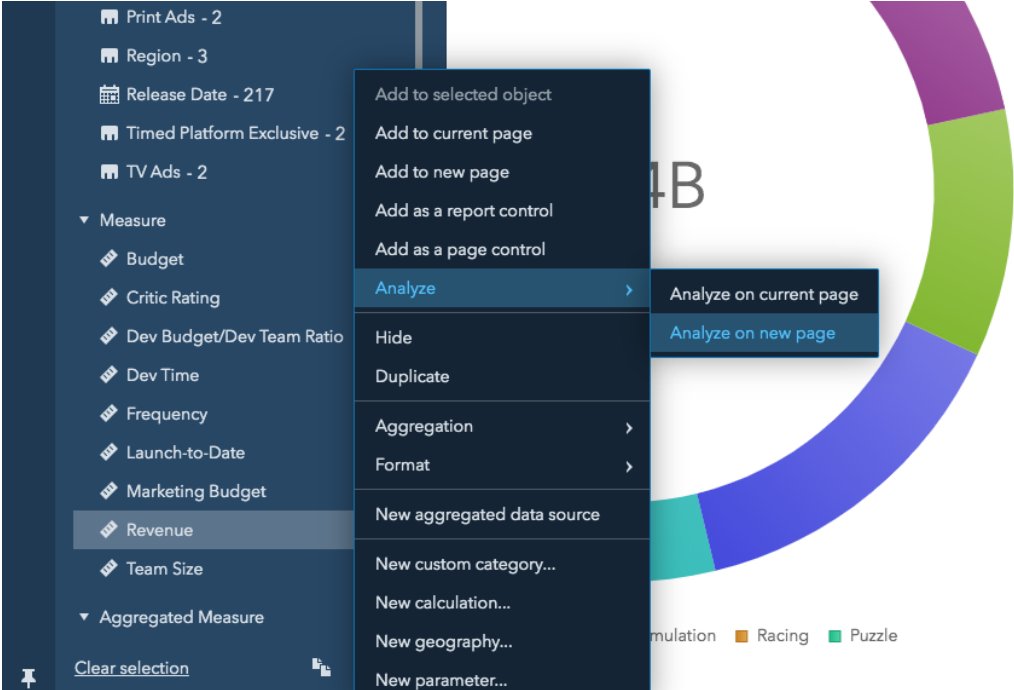


Figure 2. Select a Measure Data Item That You Want to Understand

After you choose to analyze the **Revenue** data item on a new page, SAS Visual Analytics automatically creates a model to analyze all of the rows and columns in the underlying in-memory table to determine which data items are the best predictors for **Revenue**. To understand how SAS Visual Analytics does this, you can refer to **How Does Automated Analysis Work?** on page 16. The results of the automated analysis are displayed on the new page, as illustrated in Figure 3.

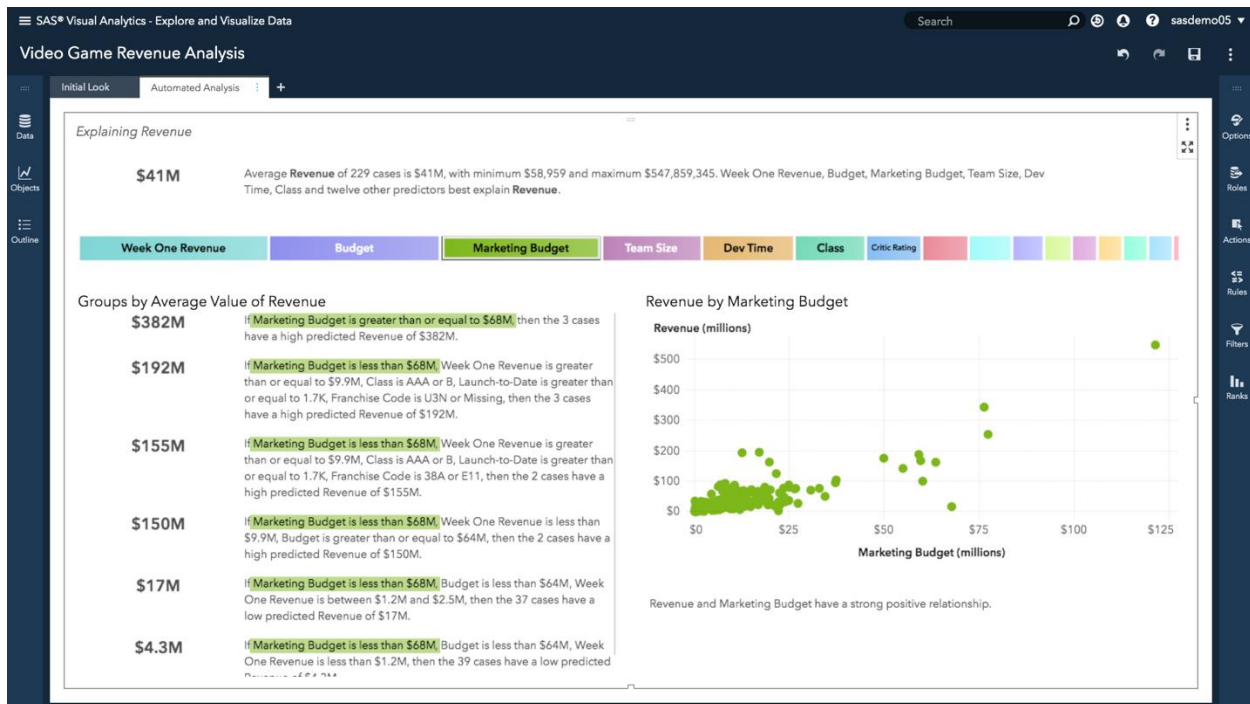


Figure 3. Automated Analysis Results That Explain the Revenue Measure Data Item

At the top of the page, you see a text-based overview of **Revenue** including the average, minimum, and maximum values, as well as a list of the most important data items that influence revenue. Figure 4 shows a closer view of this text.

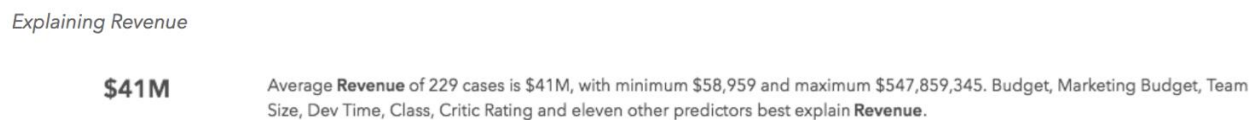


Figure 4. Natural Language Overview

Below the overview text section, you see color-coded segments on a horizontal bar, as shown in Figure 5. This horizontal bar compares the relative importance of each data item that influences **Revenue**, where the most important data item appears on the left. These data items are often referred to as underlying factors. If you mouse over each colored segment, a pop-up displays the actual importance score with one being the most important assigned to the leftmost data item segment.



Figure 5. Data Item Relative Influence on Revenue

Automated analysis in SAS Visual Analytics can assist you by eliminating data items from the analysis, such as data items having several rows with missing values. However, only you know which data items represent the factors that the organization cannot change. For example, in Figure 6, **Week One Revenue** is the most important data item predicting **Revenue**. However, **Week One Revenue** might be considered to be something that a decision maker cannot change directly. You can easily remove it from the analysis to improve the analysis explanation by focusing on data items that your organization can adjust, such as **Marketing Budget**. To remove a data item, you can simply right-click the data item on the horizontal underlying factors bar and select **Remove**.

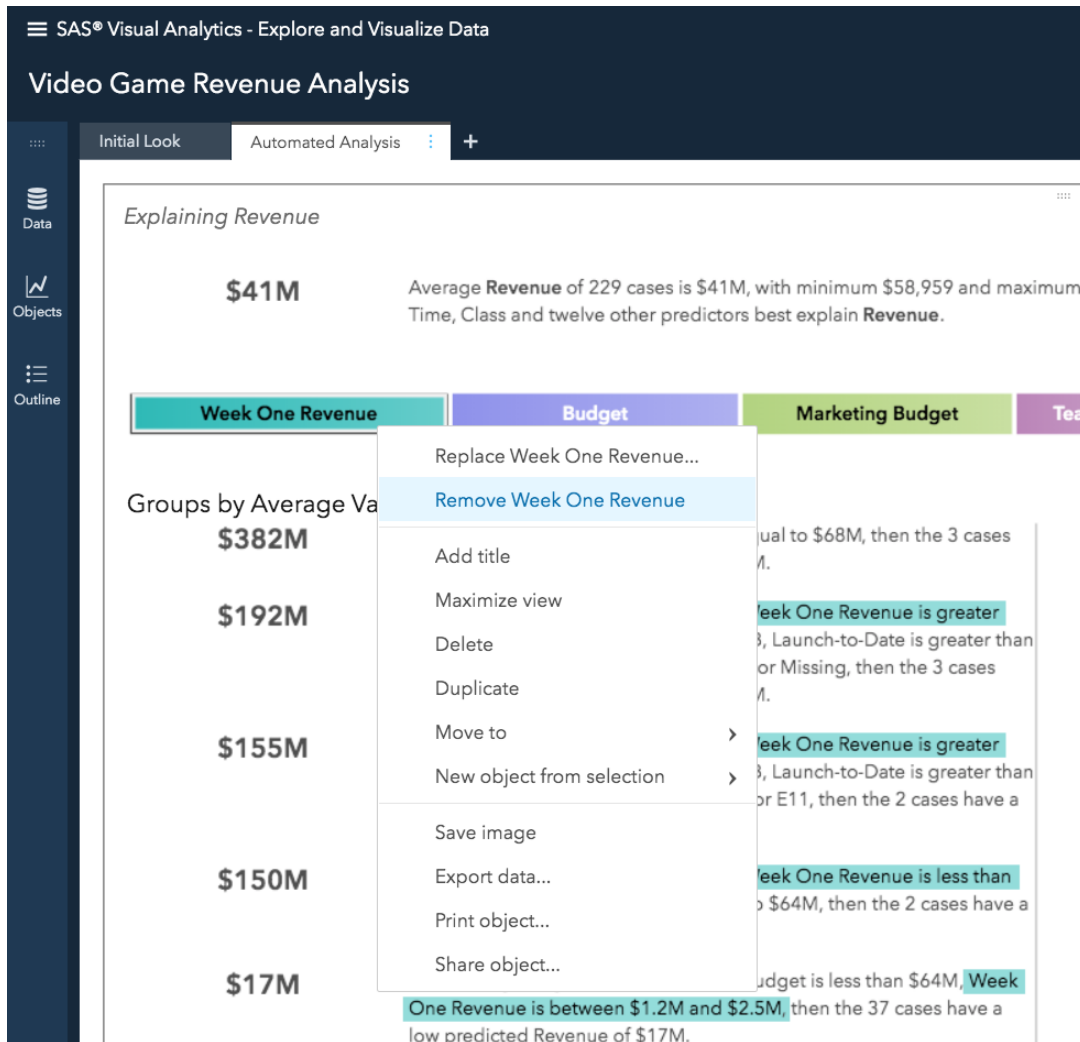


Figure 6. The Menu Option to Remove the Week One Revenue Data Item

Figure 7, illustrates how SAS Visual Analytics adapts the automated analysis results quickly after excluding the **Week One Revenue** data item from the analysis.

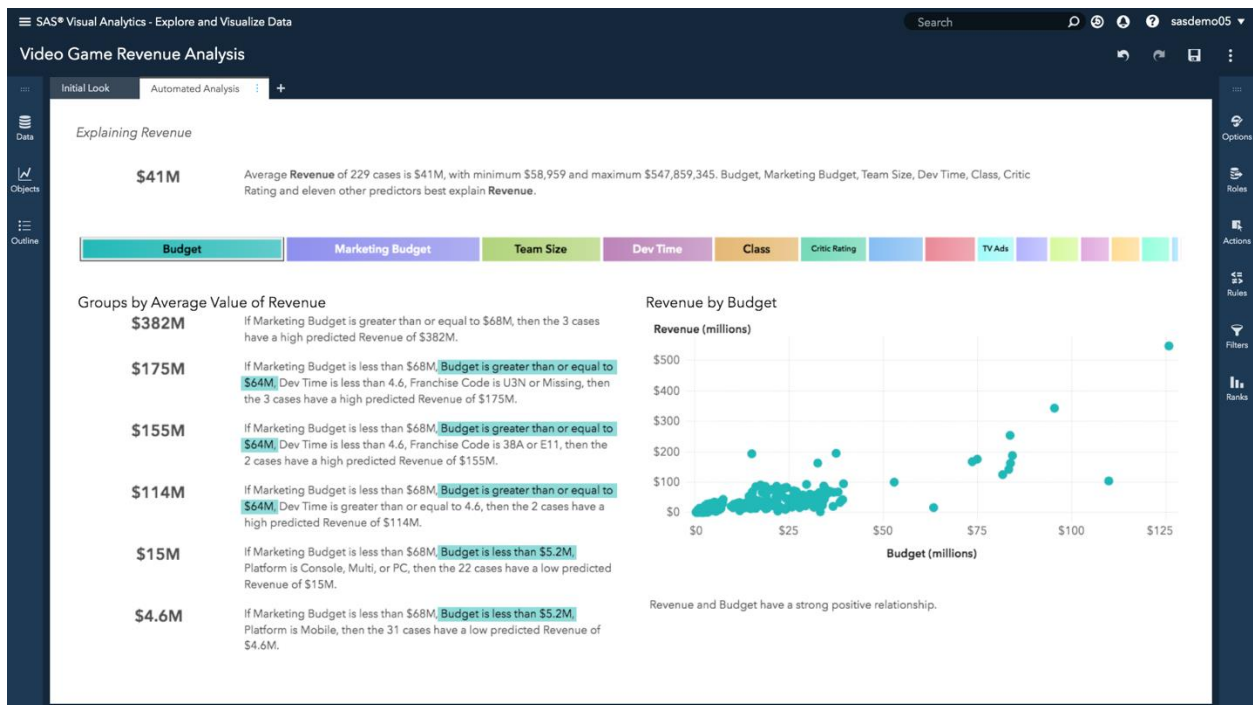


Figure 7. The Regenerated Analysis Without the Week One Revenue Data Item

Now that you know which data items most influence **Revenue**, how can you determine which data values for these important data items can lead to higher or lower revenue predictions? You would like to maximize **Revenue**, in this case, or perhaps minimize an expense metric in another scenario.

When a measure data item, like **Revenue**, is selected to be analyzed, the automated analysis results display a list of groups (or subsets) on the left of the page, below the horizontal bar. These groups further explain our video game revenue based on combinations of these influential data items with specific values. By default, the first four groups result in the greatest average values of revenue. The final two groups result in the smallest average revenue amounts and appear at the bottom of the group list. For example, you can see in Figure 8, that you have 31 games with a low predicted revenue of \$4.6 million when the game is available on the mobile platform only. You might consider recommending to your manager to add support for other platforms for these 31 games to increase **Revenue**. You can show all subgroups by selecting the **Show all subgroups** option in the **Options** pane (which is available on the far right).

Groups by Average Value of Revenue

\$382M	If Marketing Budget is greater than or equal to \$68M, then the 3 cases have a high predicted Revenue of \$382M.
\$175M	If Marketing Budget is less than \$68M, Budget is greater than or equal to \$64M, Dev Time is less than 4.6, Franchise Code is U3N or Missing, then the 3 cases have a high predicted Revenue of \$175M.
\$155M	If Marketing Budget is less than \$68M, Budget is greater than or equal to \$64M, Dev Time is less than 4.6, Franchise Code is 38A or E11, then the 2 cases have a high predicted Revenue of \$155M.
\$114M	If Marketing Budget is less than \$68M, Budget is greater than or equal to \$64M, Dev Time is greater than or equal to 4.6, then the 2 cases have a high predicted Revenue of \$114M.
\$15M	If Marketing Budget is less than \$68M, Budget is less than \$5.2M, Platform is Console, Multi, or PC, then the 22 cases have a low predicted Revenue of \$15M.
\$4.6M	If Marketing Budget is less than \$68M, Budget is less than \$5.2M, Platform is Mobile, then the 31 cases have a low predicted Revenue of \$4.6M.

Figure 8. Groups by Average Value Further Explain the Revenue Data Item

The visualization on the right and below the horizontal factors bar shows the relationship between the **Revenue** data item and the underlying factor data item selected in the horizontal bar. As you click each data item segment in the horizontal factors bar, the visualization changes in the bottom right. The contents of this relationship visualization depend on the data item types of both the selected data item in the Data pane for automated analysis and the selected underlying factor data item on the horizontal bar. Table 1 is based on a similar table from the SAS Visual Analytics 8.3 online documentation (SAS Institute Inc., 2018).

Data Item Type Selected for Analysis	Measure Underlying Factor Data Item Selected	Category Underlying Factor Data Item Selected
Measure Data Item Selected	A scatter plot or heat map	A bar chart
Category Data Item Selected	A histogram that is grouped by two category values, the selected category data item value and one all other value	A stacked bar chart that is grouped by two category values, the selected category data item value and one all other value

Table 1. The Visualization Depends on the Types of Both of the Data Items Selected for Analysis and the Data Item Selected on the Underlying Factors Bar

AUTOMATED ANALYSIS FOR A CATEGORY

Outcomes in your data are often represented as numerical values, known as metrics. Though, automated analysis in SAS Visual Analytics isn't just limited to analyzing numerical performance. Outcomes can be represented as descriptive categories. For example, the status of out-of-stock might represent an outcome of interest for an online retailer. For a category data item, you can select one from the Data pane on the left of the screen for automated analysis in the same way that you could choose the **Revenue** data item in the previous example. Figure 9 illustrates the selection of a category data item for **TV Ads**.

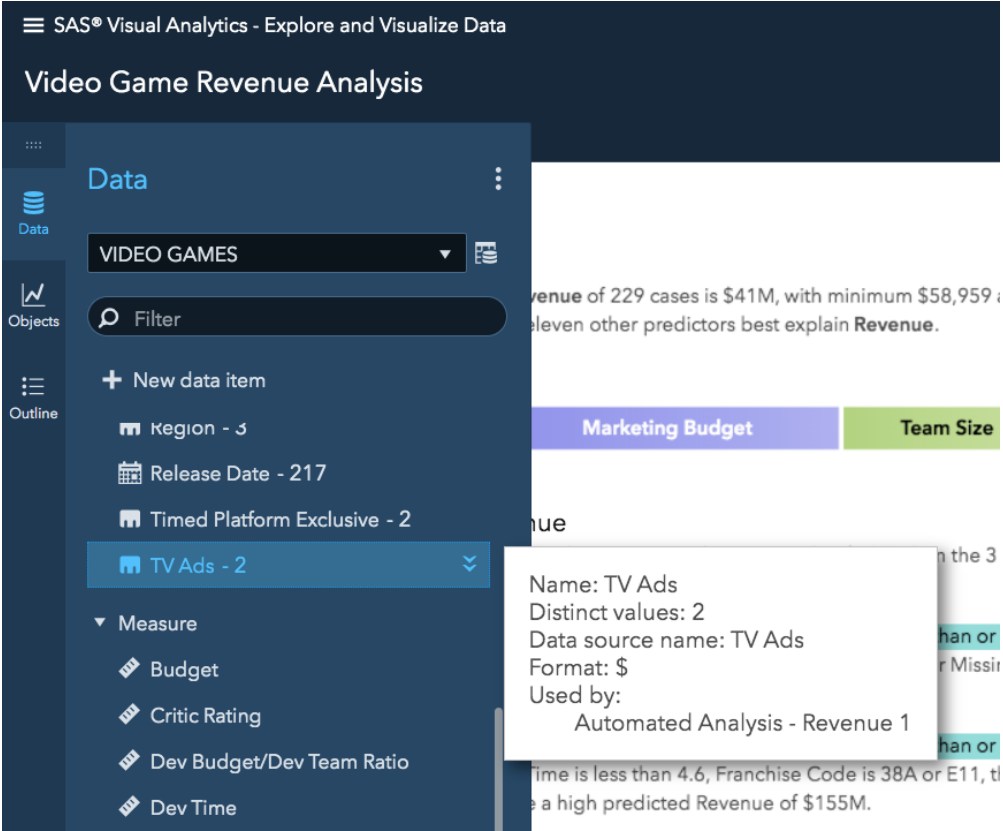


Figure 9. TV Ads is a Category Data Item

As shown in Figure 10, You can right-click the category data item that you are interested in and select **Analyze**.

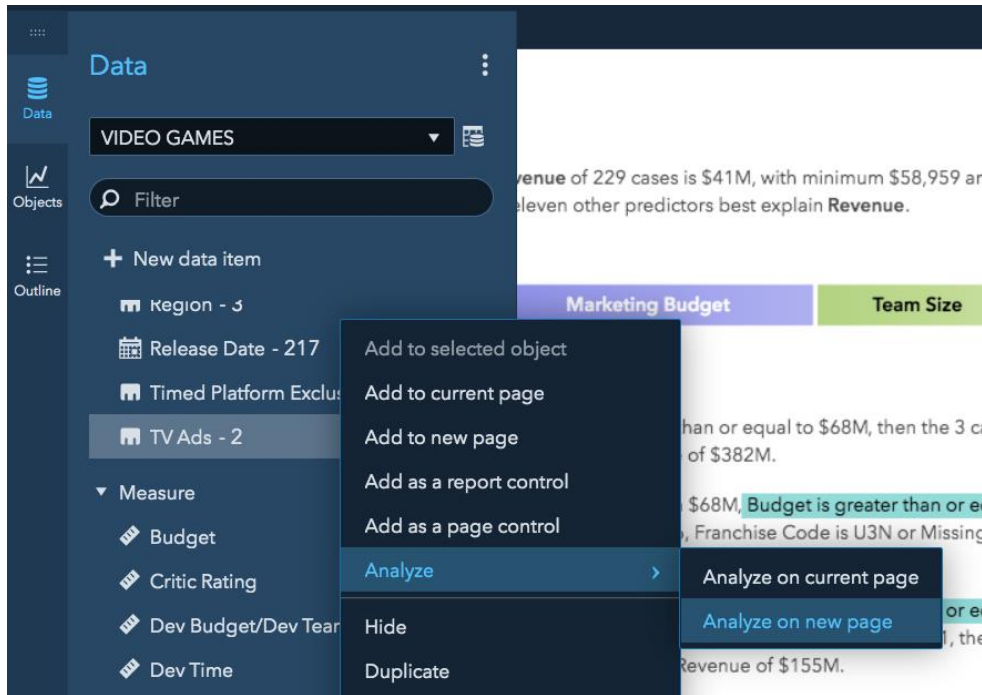


Figure 10. Select the TV Ads Category Data Item to Analyze on a New Page

As with the **Revenue** data item before, SAS Visual Analytics processes all of the rows and columns to determine which data items are underlying factors influencing whether a video game has TV ads or not. Then, it generates groupings, visualizations, and natural language explanations to help non-technical viewers understand what is determining which video games are associated with **TV ads**.

As you can see near the top and to the right of the page in Figure 11, you can select the value for the category data item to be analyzed. In the case of the **TV Ads** data item, the choice is limited to either **Yes** or **No**.

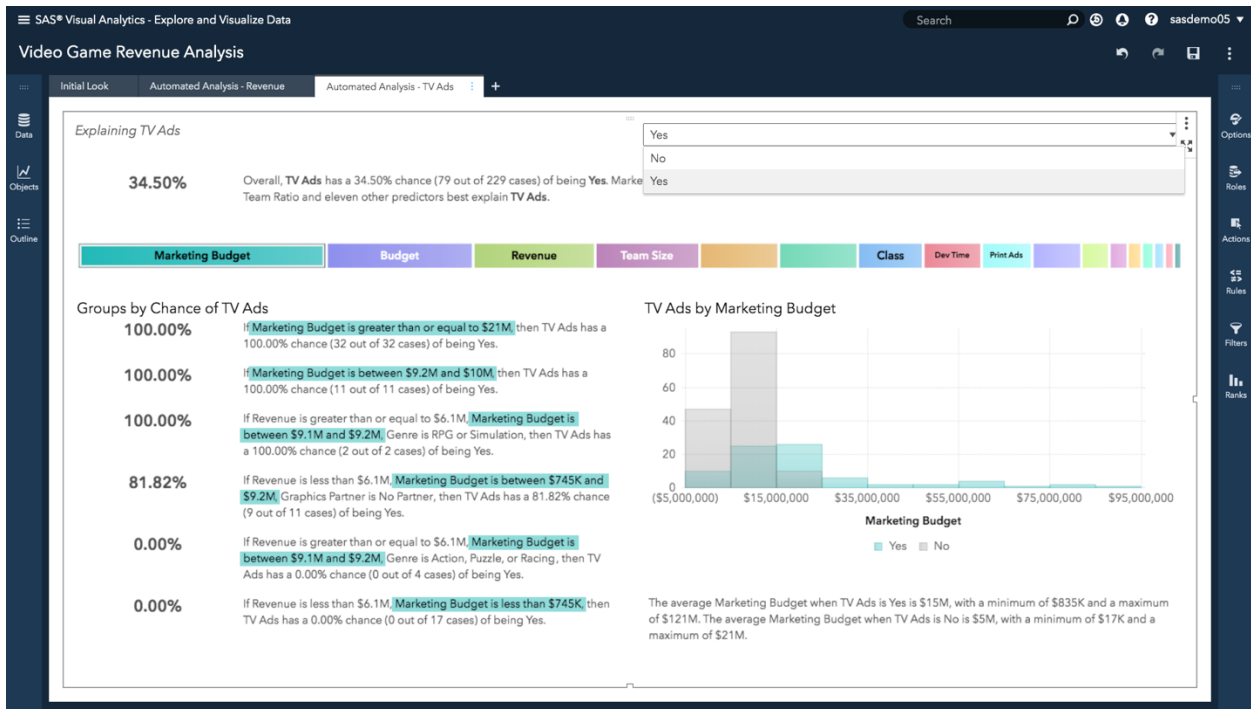


Figure 11. Change the Value of TV Ads From No to Yes

For category data items, by default, the analysis results, in the bottom left of the page, first displays the four groups that contain the greatest percentages of the response **Yes** for **TV Ads**. Likewise, the list of groups also displays the two groups that contain the least percentages of the response **Yes** at the bottom of the page.

As stated in Table 1 on page 8, the visualization displayed on the bottom right of the page is a histogram that is grouped by two category values, the selected category data item value, and an all other value, when a measure data item is selected on the horizontal underlying factors bar. For the visualization on the right in Figure 11, shown above, the underlying factor data item is **Marketing Budget** and the all other value is simply **No** since there are only two values for the **TV Ads** category data item. If more than two values exist for the category data item being analyzed, the gray portion of the histogram bars would be represented in the legend as **Not** plus the selected categorical value. If the underlying factor data item selected on the horizontal bar (near the top of the page) is a category instead of a measure, a bar chart that is grouped by two character values displays at the bottom right of the page instead of a histogram.

DERIVE NEW DATA ITEMS AND VISUALIZATIONS

With automated analysis in SAS Visual Analytics, you can easily derive new data items and new visualizations from the automated analysis results in order to enhance your analysis and story. A new visualization can be inserted on a new page to compare the portion of your data that is included to the portion that is excluded from one of the subgroups on the analysis results page. In addition, you can create a derived data item based on one of the subgroups from the automated analysis results. You could use the derived data item to filter visualizations in a new or existing dashboard, for example. The first step to derive a new visualization or new data item is to click a subgroup in the automated explanations on the bottom left of the page, as show in Figure 12.

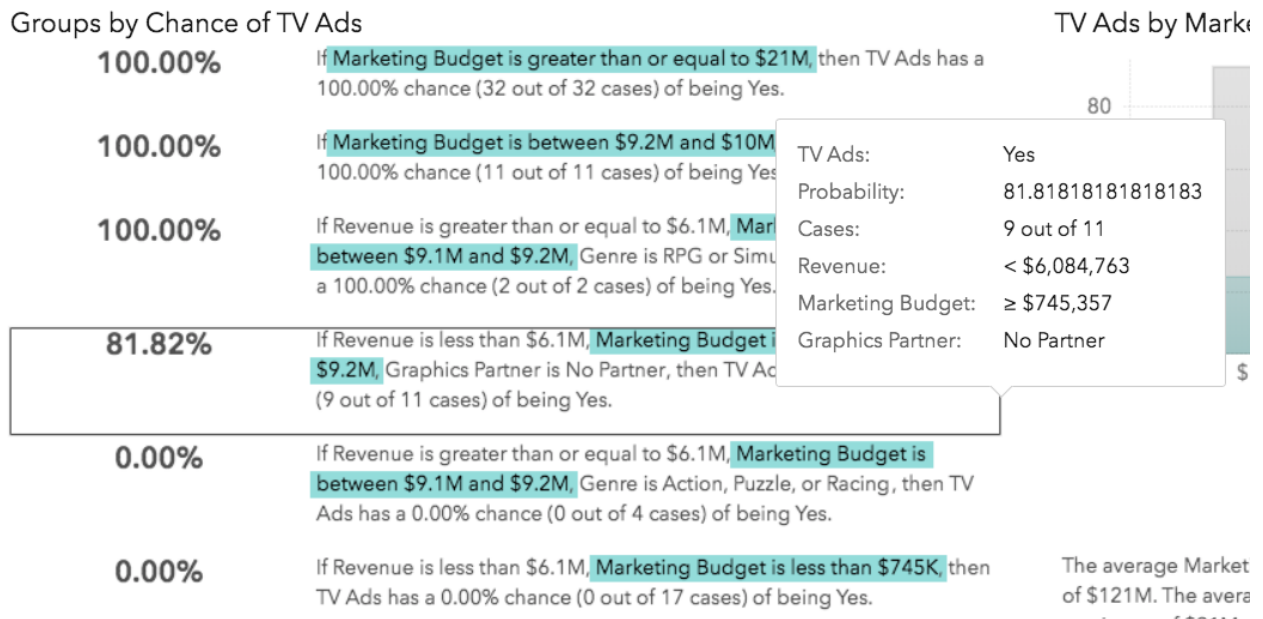


Figure 12. Select a Subgroup From the List

To create a new visualization, right-click the subgroup and select **New object from subgroup on new page**. Then, select a data item that you want to include in the visualization, as shown in Figure 13.

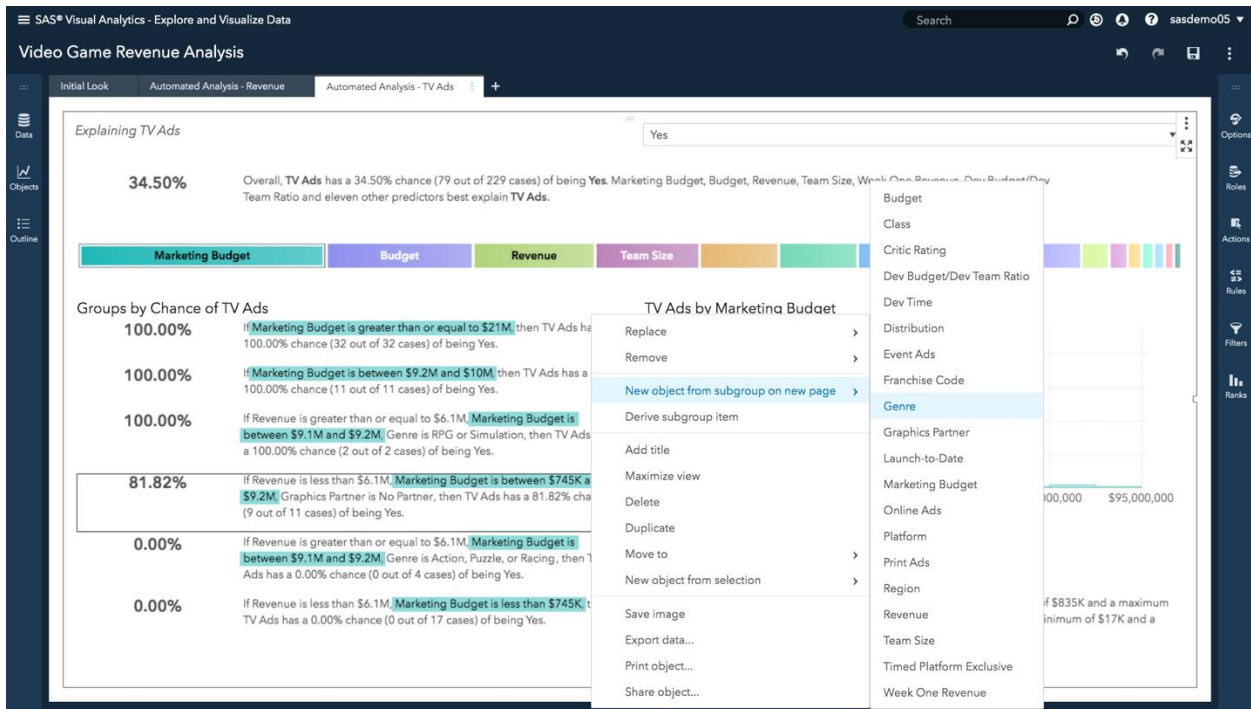


Figure 13. Create a New Visualization From the Selected Subgroup (on the Left)

A new page displays the new visualization, and a new data item is created for the selected subgroup automatically. The visualization in Figure 14, shows the portion of the **TV Ads** category data item values that is included in (blue bars) and the portion excluded from (purple bars) the 81.82% subgroup.

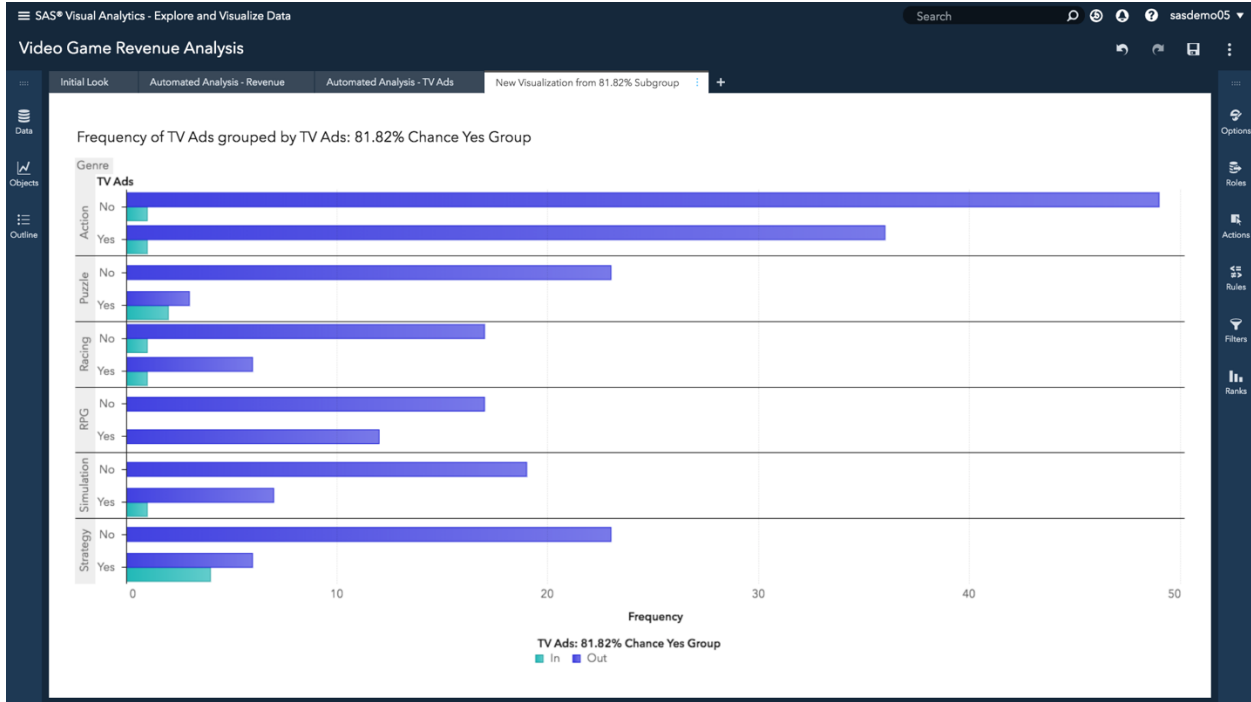


Figure 14. Compare the Frequency by Genre For Those In and Those Out of the 81.82% Subgroup

When you create a new object from a subgroup, a new data item representing the selected subgroup is automatically created. To derive a subgroup data item manually, right-click the subgroup, and then select **Derive subgroup item**. With either approach, a new category data item is created in the Data pane, as show in Figure 15. You can use this new derived data item just as you would any other data item available in the Data pane.

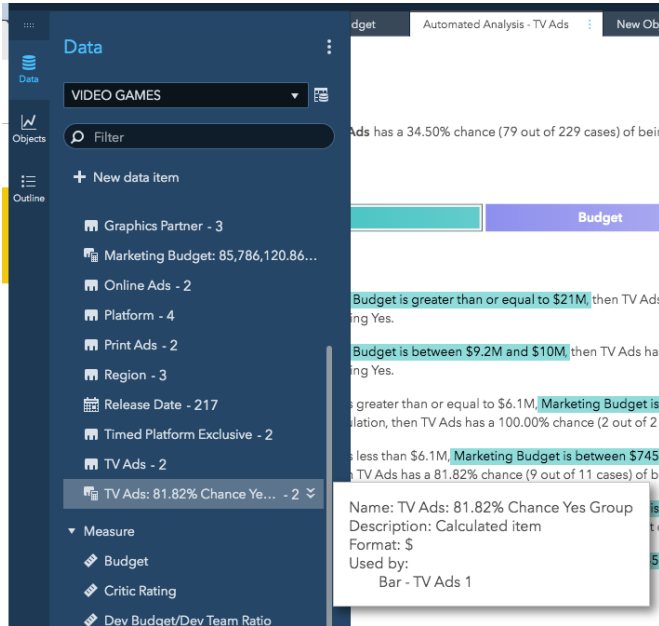


Figure 15. A New Calculated Data Item Appears in the Data Pane for the Selected Subgroup

You can see and modify, the derived subgroup data item expression if you right-click the new data item, and then select **Edit**, as shown in Figure 16.

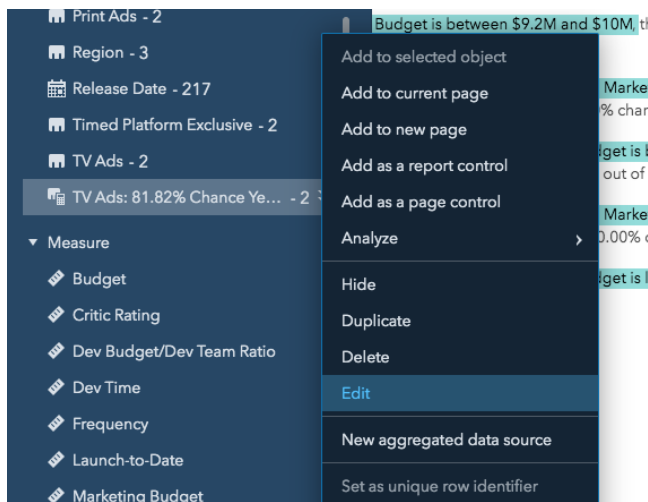


Figure 16. Right-Click and Select Edit to See the Subgroup Condition Expression

Figure 17 illustrates the IF ELSE expression that was created automatically when the new data item was derived from the automated analysis subgroup. You could create a custom subgroup manually by tweaking any of the values in this expression.

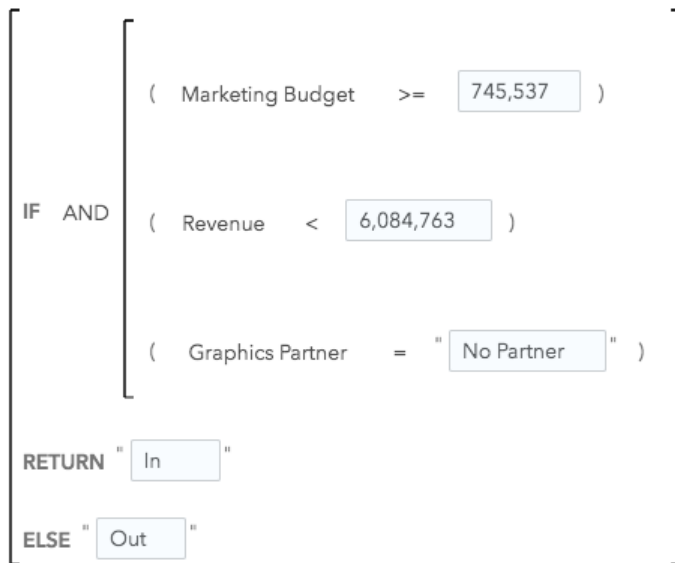


Figure 17. Conditional Expression for the Automatically Created Subgroup Data Item

HOW DOES AUTOMATED ANALYSIS WORK?

You don't need to know how SAS Visual Analytics automates the analysis process in order to use it. That's the point of automated analysis. It allows a wide range of non-technical business users to take advantage of the insights. However, if you're curious how it works, this section explains the underlying process. The rules used to determine the relationship visualization displayed on the bottom right of the page are provided in Table 1 on page 8. In the SAS Visual Analytics 8.3 release, you cannot modify the rules or modeling techniques used to create the insights. The underlying process that SAS Visual Analytics 8.3 uses to generate automated analysis results is subject to change as SAS enhances the capability in future releases.

After you choose to analyze a data item for the outcome you are interested in, most of the remaining data items are added as underlying factors for the analysis. Underlying factor data items are pruned or excluded using the following set of rules:

- ID or ID-like data items are excluded.
- Duplicated data items are excluded.
- Identical items to the data item of interest (for example, 1 or 0 versus Yes or No) are excluded.
- High percentage missing value data items are excluded.
- High cardinality data items are excluded.

After SAS Visual Analytics prunes the underlying factor data items using the rules above, an optimal one-level decision tree is created to split the data for the remaining underlying factors. The factors are ranked based on the decision tree. SAS Visual Analytics does this by

creating 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.

To view the screening result action taken and relative importance scores for each data item, you can maximize the automated analysis object using the maximize icon in the top right just under the object's menu icon. Refer to the table that appears below the normal automated analysis results. This table appears only when the object is maximized.

To create the high and low subgroups in the bottom left of the automated analysis results, SAS Visual Analytics uses a decision tree again. A decision tree is a recursive partition method that divides data into multiple subsets. It is widely used to identify subgroups with high or low values for a variable of interest.

The natural language text, displayed at the top overview of the automated analysis results, in the subgroup listing, and below the relationship visualization, explains insights in terms most business users will understand. To generate this text, SAS Visual Analytics uses Apache Velocity templates to define text, styling, variable placement, and logic. It sends one of these templates along with data to a SAS Natural Language Generation (NLG) microservice to generate each natural language explanation displayed in the results. SAS Visual Analytics 8.3 supports only English for generated text.

CONCLUSION

There is more data than ever before, but there isn't more time. Data scientists are extremely valuable, but their availability can be quite scarce. Dramatic increases in compute power, reduction in infrastructure costs, reduced time to deploy software using cloud technologies, and Natural Language Processing (NLP) are just some of the leaps forward in recent years that have made now a good time to automate the manual steps required to get deeper insights from your data. Beginning in the September 2018 update, SAS made a new automated analysis feature available in SAS Visual Analytics 8.3. If you already have SAS Visual Analytics 8.3 with the September 2018 update, I'd recommend that you give automated analysis a quick try. Then, share your findings with a few of your close peers. Next, pilot automated analysis for a specific organizational outcome and include a non-technical business user that is interested in this outcome. If you do not have SAS Visual Analytics 8.3 or higher yet, I encourage you to sign up for a [free trial of SAS Visual Analytics](#) in order to play with the automated analysis capability using some of your data.

REFERENCES

SAS Institute Inc. 2018. *SAS Visual Analytics 8.3: Working with Automated Analysis*. Cary, NC: SAS Institute Inc. Available <http://go.documentation.sas.com/?cdcId=vacdc&cdcVersion=8.3&docsetId=vaobj&docsetTarget=n0ewwfd6udhv7qn1nropd18lof9j.htm>

Diamandis, Peter and Steven Kotler. 2012. *Abundance: The Future Is Better Than You Think*. New York, NY: Free Press.

ACKNOWLEDGMENTS

The author would like to thank four SAS colleagues for their help and prior work that contributed to this paper. Don Chapman, the R&D development director responsible for developing the automated analysis capability answered many questions and provided examples. XiangXiang Meng, the product manager who defined the initial requirements for the automated analysis feature, co-presented this topic with Don at Analytics Experience 2018. Their presentation materials were quite helpful. Atrin Assa published a brief YouTube video on automated analysis in December 2018. The data and scenario Atrin demonstrated was adapted for this paper. Finally, Melanie Carey published an informative three-part blog series on the automated analysis feature in SAS Visual Analytics 8.3 in January 2019. The details and explanations she provided helped shape the content of this paper. See the Recommended Reading section in this paper for links to Atrin and Melanie's resources.

RECOMMENDED READING

- Carey, Melanie. "How SAS Visual Analytics' automated analysis takes customer care to the next level – Part 1" Available <https://blogs.sas.com/content/sgf/2019/01/02/how-sas-visual-analytics-automated-analysis-takes-customer-care-to-the-next-level-part-1>. Last modified January 2, 2019. Accessed on March 28, 2019.
- Carey, Melanie. "How SAS Visual Analytics' automated analysis takes customer care to the next level – Part 2" Available <https://blogs.sas.com/content/sgf/2019/01/07/how-sas-visual-analytics-automated-analysis-takes-customer-care-to-the-next-level>. Last modified January 7, 2019. Accessed on March 28, 2019.
- Carey, Melanie. "How SAS Visual Analytics' automated analysis takes customer care to the next level – Part 3" Available <https://blogs.sas.com/content/sgf/2019/01/14/how-sas-visual-analytics-automated-analysis-takes-customer-care-to-the-next-level-2>. Last modified January 14, 2019. Accessed on March 28, 2019.
- Assa, Atrin. "Automated Analysis with SAS Visual Analytics" Available <https://www.youtube.com/watch?v=Jl-zZ7LhXfE>. Available December 19, 2018. Accessed on March 28, 2019.
- Carey, Melanie. "Using the Automated Analysis Feature in SAS Visual Analytics in SAS Viya" Available <https://video.sas.com/detail/video/5978210714001/using-the-automated-analysis-feature-in-sas-visual-analytics-in-sas-viya>. Available December 11, 2018. Accessed March 28, 2019.
- Brown, Anna. "Automated Analysis What's the deal with AI-powered automated analysis in SAS Visual Analytics?" Available <https://communities.sas.com/t5/SAS-Visual-Analytics/What-s-the-deal-with-AI-powered-automated-analysis-in-SAS-Visual/td-p/539716>. Last modified March 1, 2019. Accessed on March 28, 2019.

CONTACT INFORMATION

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

Rick Styll
SAS Institute Inc.
rick.styll@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.