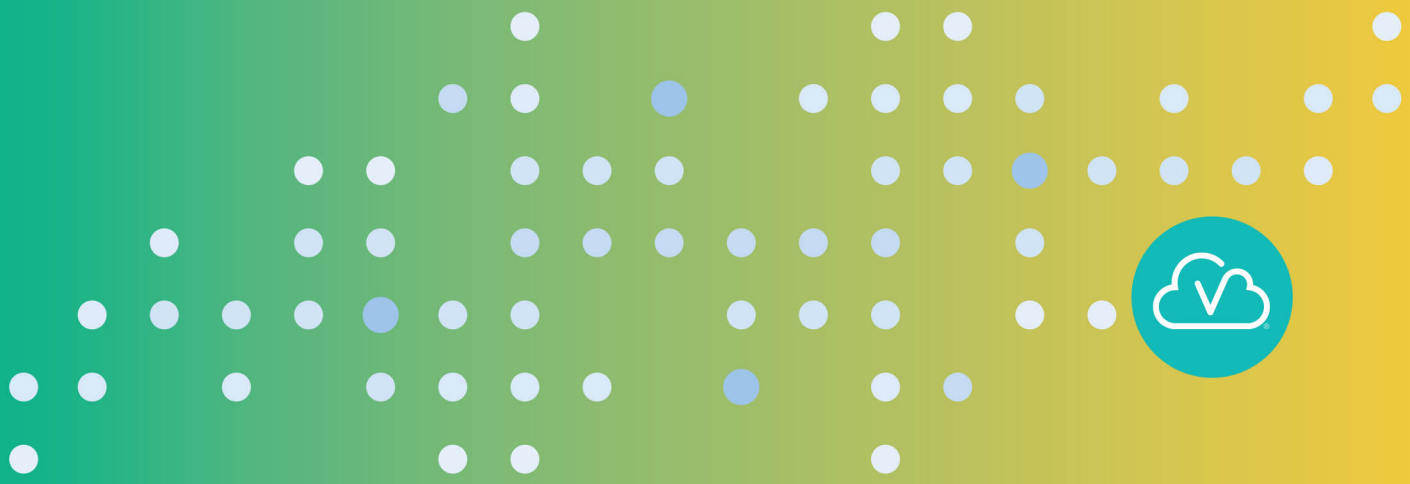
Exploring SAS® Viya®

Visual Analytics, Statistics, and Investigations



The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2019. *Exploring SAS® Viya®: Visual Analytics, Statistics, and Investigations*. Cary, NC: SAS Institute Inc.

Exploring SAS® Viya®: Visual Analytics, Statistics, and Investigations

Copyright © 2019, SAS Institute Inc., Cary, NC, USA

978-1-64295-490-6 (Paperback)
978-1-64295-453-1 (Web PDF)

All Rights Reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

June 2019

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.

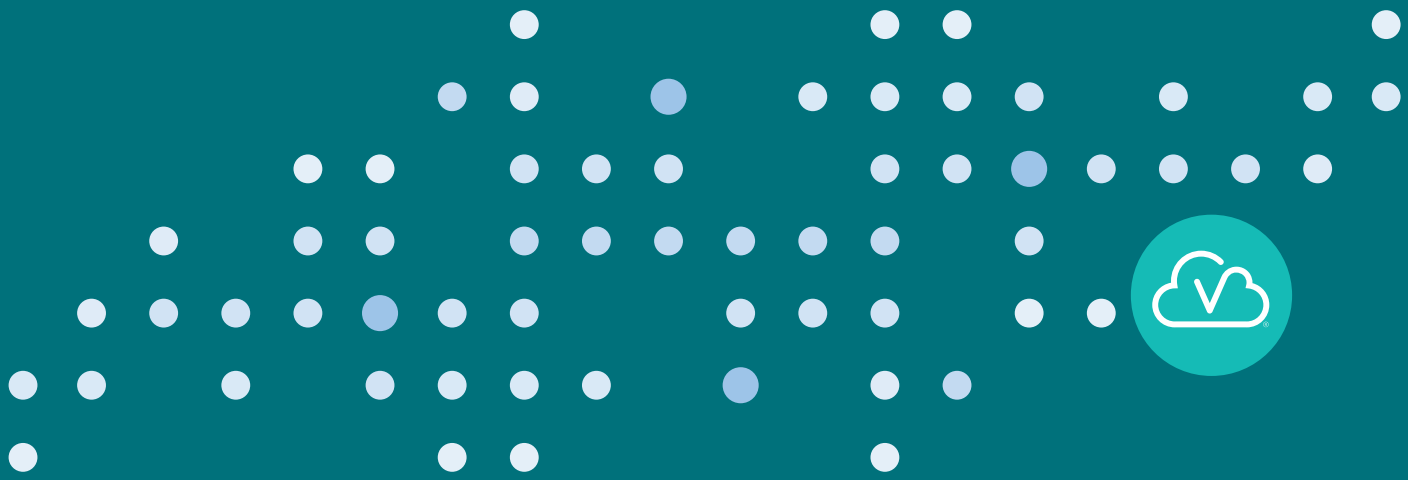
SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to **http://support.sas.com/thirdpartylicenses**.

Contents

Chapter 1: Viewing and Creating Reports in SAS Visual Analytics 1
Customizing Your SAS Home Page 1
Using SAS Report Viewer 7
Creating a Basic Report 13
Creating an Advanced Report with a Forecast 21
Modifying Reports 31
Resources 42
Chapter 2: SAS Visual Statistics 43
Introduction 43
Getting Started with SAS Visual Statistics 43
Clustering Model 45
Decision Tree Model 49
Logistic Regression Model 54
Resources 59
Chapter 3: SAS Visual Text Analytics 61
Introduction 61
Batch Score a Topic Model 61
Create Custom Concepts 67
Explore Category Output Data Interactively 71
Resources 79
Chapter 4: SAS Visual Investigator 81
Introduction 81
Searches 85
Alert Triage Process 94
Workflow Process 96
Export to Microsoft Excel 98
Structured Printing 100
Audit Reports for Administrators 100
Resources 102

Free SAS® Viya® e-Books: Fundamentals

This series is based on content from SAS® Viya® Enablement, a free course available from SAS Education. You can follow along with the examples in real time by watching the videos if you prefer. Topics covered illustrate the features and capabilities of SAS Viya. SAS Viya extends the SAS platform to enable everyone – data scientists, business analysts, developers, and executives alike – to collaborate and realize innovative results faster.



support.sas.com/freesasebooks

sas.com/books
for additional books and resources.

sas
THE POWER TO KNOW®

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. © 2019 SAS Institute Inc. All rights reserved. M1913158 US.0419

About This Book

What Does This Book Cover?

Data visualization is the presentation of data in a pictorial or graphical format. It enables decision makers to see analytics presented visually so that they can grasp difficult concepts or identify new patterns. With interactive visualizations, you can take a concept further by using technology to navigate through charts and graphs for more detail, interactively changing what data you see and how it is processed. SAS offers several solutions for visualizing your data, many of which are powered by SAS® Viya®.

SAS Viya is an open analytics platform that can handle any data type, volume, or speed. A cloud-enabled, in-memory analytics engine, it is elastic, scalable, and fault tolerant. It contains a standardized code base that supports programming in SAS and other languages, such as Python, R, Java and Lua. In addition, it can deploy seamlessly to any infrastructure or application ecosystem with support for cloud, on-site, or hybrid environment. The high-performance processing power of SAS Viya is provided by SAS Cloud Analytics Services (CAS). CAS is an in-memory engine that can dramatically accelerate data management and analytics with SAS.

The four visualization solutions powered by SAS Viya that will be discussed in this book are as follows:

- SAS Visual Analytics
- SAS Visual Statistics
- SAS Visual Text Analytics
- SAS Visual Investigator

Users of all skill levels can visually explore data on their own in these programs while drawing on powerful in-memory technologies for faster analytic computations and discoveries. These programs offer an easy-to-use self-service environment that can scale on an enterprise-wide level.

The content in this book is based on SAS® Viya® Enablement, a free course available from SAS Education. This book covers how to view, create, and manipulate reports in SAS Visual Analytics and SAS Model Studio environments. This book only begins to show what these programs can do. It mainly focuses on the point-and-click features of the software.

If you want to learn more about the features of SAS Viya, how to load data into the CAS server, how to write new code, and how to perform data management and administrative tasks, then you might be interested in reading *Exploring SAS® Viya®: Programming and Data Management*.

Is This Book for You?

SAS visualization software is designed for anyone in your organization who wants to use and derive insights from data—from influencers, decision makers, and analysts to statisticians and data scientists. It also offers IT an easy way to protect and manage data integrity and security.

With web-based exploratory analysis and other easy-to-use features, even users without analytical expertise can use predictive analytics to gain precise insights. Nontechnical users can create and change queries simply by selecting items from a sidebar or dynamically filtering and grouping data items. Autocharting selects the visualization that is most suitable for the type of data chosen. “What does it mean” pop-up boxes provide explanations of analytical techniques, helping everyone understand the data and what the analysis means.

What Should You Know about the Examples?

The content in this book is based on SAS® Viya® Enablement, a free course available from SAS Education. You can follow along with the examples in real time by watching the videos if you prefer.

This book includes tutorials for you to follow to gain hands-on experience with SAS Visual Analytics and SAS Model Studio. Wherever possible, the source of the sample data is provided in a link. Some features shown might be available only if your site has licensed that feature in SAS Viya. Therefore, the options in your version of SAS might look different.

We Want to Hear from You

Do you have questions about a SAS Press book that you are reading? Contact us at saspress@sas.com.

SAS Press books are written *by* SAS Users *for* SAS Users. Please visit sas.com/books to sign up to request information about how to become a SAS Press author.

Learn about new books and exclusive discounts. Sign up for our new books mailing list today at https://support.sas.com/en/books/subscribe-books.html.

Chapter 1: Viewing and Creating Reports in SAS Visual Analytics

Customizing Your SAS Home Page 1
- Welcome Banner 2
- Menu Bar 3
- Application Bar Settings 6
- Help 7

Using SAS Report Viewer 7
- Slider 9
- Filter 9
- Details 10
- Clear Filters 11
- Information 11
- Close Report 12

Creating a Basic Report 13
- Data Panel 14
- Objects Panel 15
- Add a Page 18
- Connecting Pages 20

Creating an Advanced Report with a Forecast 21
- Modifying Time Values 21
- Modifying Measures 23
- Assigning Items to the Report 24
- Adding a Forecast 25
- Explore Mode 26

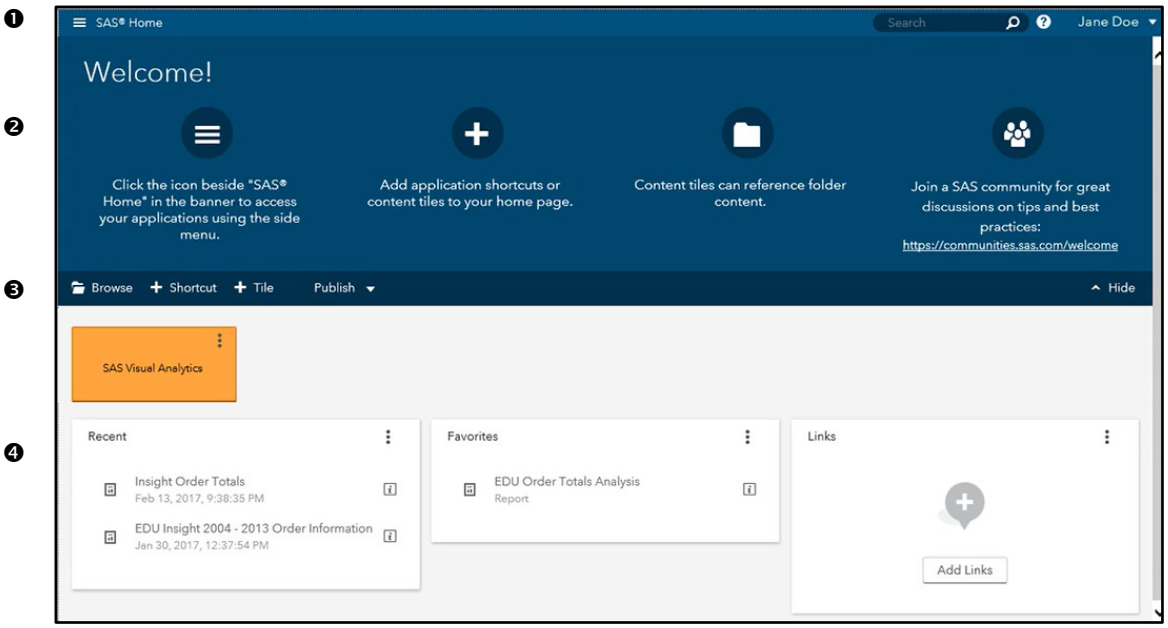
Modifying Reports 31
- Using Hidden Pages 33
- Rearranging Objects by Using the Outline Pane 35
- Creating New Charts Using Duplication and Changing Chart Types 38

Resources 42

Customizing Your SAS Home Page

Let's explore the features of the Home page in SAS Visual Analytics and learn how to make some customizations. After you sign in, you can see that the SAS Home Page consists of several sections as shown in Figure 1.1.

Figure 1.1: Home Page



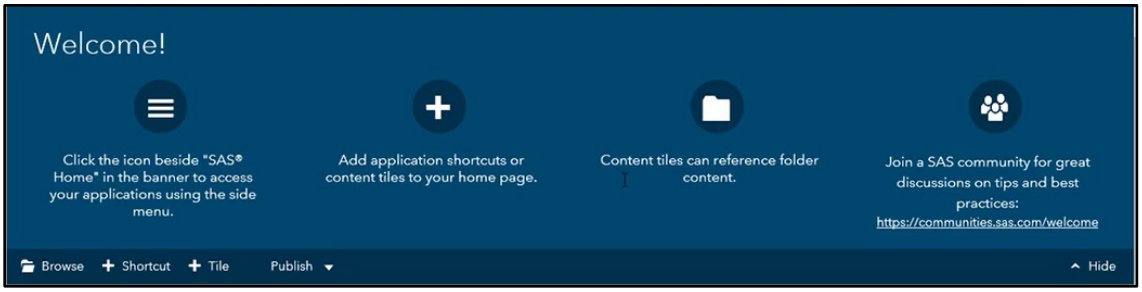
From top to bottom, we see the following:

- ❶ the **Application Bar**, which includes your sign-in name, help button, and search bar
- ❷ the **Welcome Banner**, which can be hidden to allow for more room for tiles
- ❸ the **Menu Bar**, which includes Browse, Shortcut, and Tile options
- ❹ shortcut tiles and content tiles

Welcome Banner

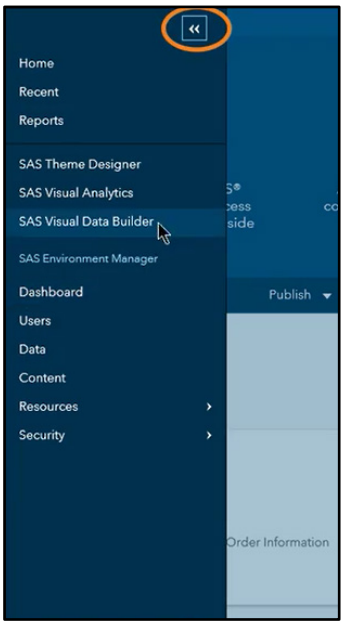
Let's start by taking a quick look at the information in the Welcome banner in Figure 1.2, which consists of four tips to help you get started.

Figure 1.2: Welcome Banner



The first tip directs us to the side menu. Clicking the icon with three horizontal lines to the left of the SAS Home in the banner expands this menu. The side menu contains links to all the SAS applications that you can use. Close the side menu by clicking the collapse icon in the upper right-hand corner (circled in orange in Figure 1.3) or by clicking anywhere outside the menu.

Figure 1.3: Side Menu



Next, there is a tip about adding shortcuts to the SAS applications that you use most frequently and about adding content tiles—for example, a tile that lists the contents of a folder where you store reports.

Finally, in the Welcome banner, you see a link to join a SAS community. These communities are helpful sites where you can continue to learn about SAS applications, see how other members of the community are using the SAS applications, or reach out to your peers with a question.

You can hide the Welcome banner by selecting Hide in the lower right corner of the banner, and you can toggle it open again by selecting Show.

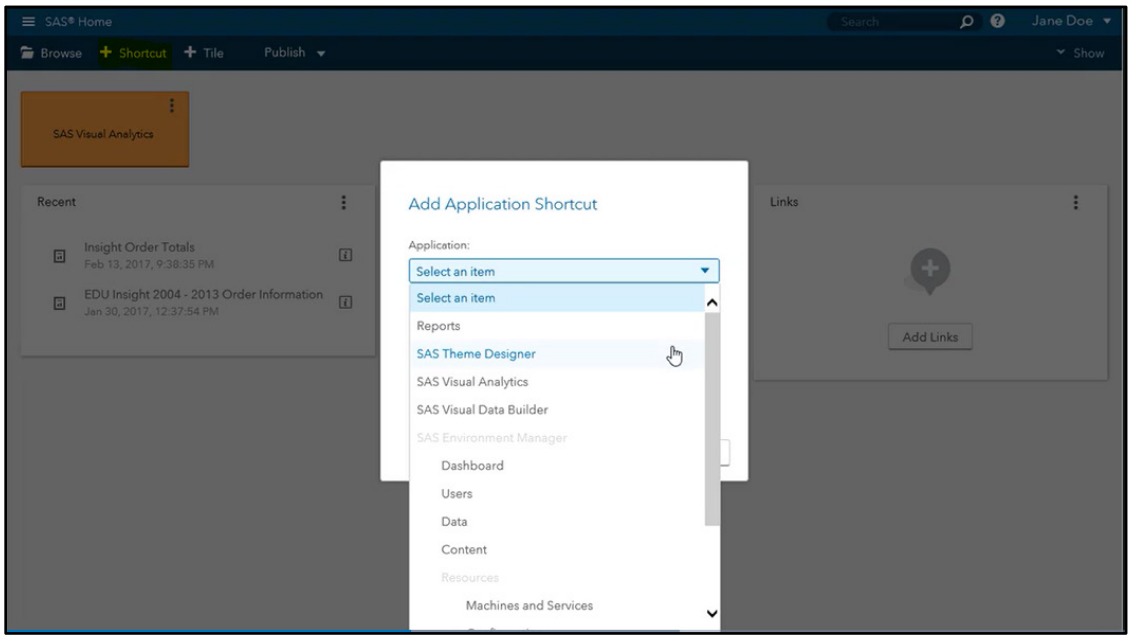
Menu Bar

Now let's make some customizations to the Home page by using the Menu Bar.

Add an Application Shortcut

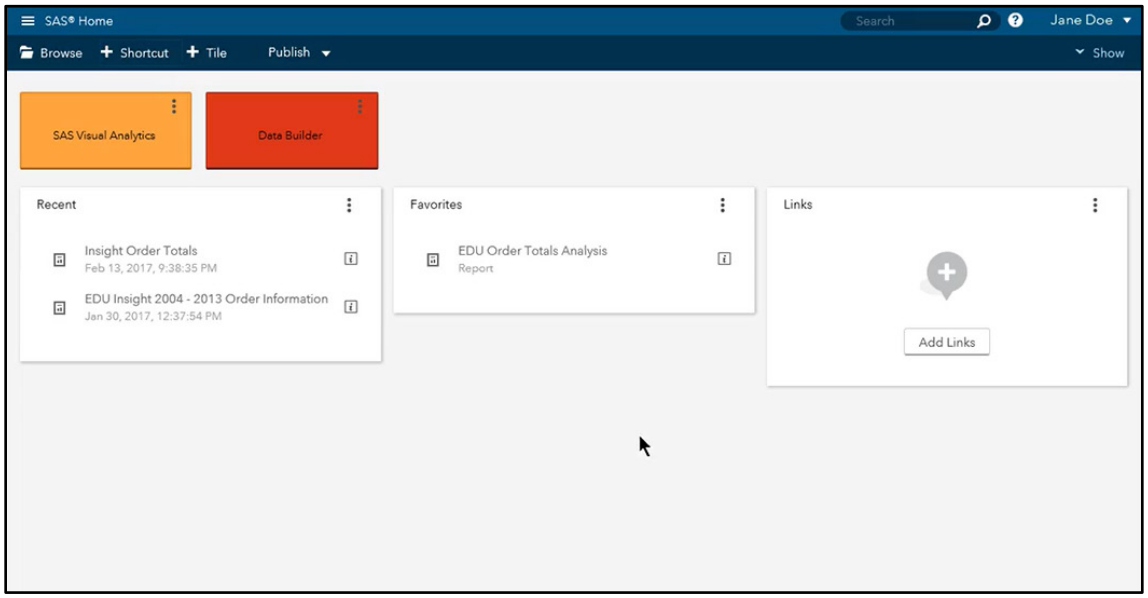
Click the Add Shortcut button and add an application shortcut to your home page. In the Add Application Shortcut window, select the application that you want to add from the drop-down list as shown in Figure 1.4. This is the same list of applications that you saw on the side menu.

Figure 1.4: Add Application Shortcut



In this example, we will select SAS Visual Data Builder. For the shortcut name, we will type "Data Builder" and select the color red. Click Save and the application shortcut is now added to your home page as shown in Figure 1.5.

Figure 1.5: Data Builder Shortcut



If you want to make a change to the application shortcut later, you can click the three vertical dots in the upper right-hand corner of each application, which opens the Options menu. You can select Edit to modify the settings or select Remove to remove the shortcut from the Home page. Notice that each tile has the options menu. The options vary by tile, but all tiles include the Hide option.

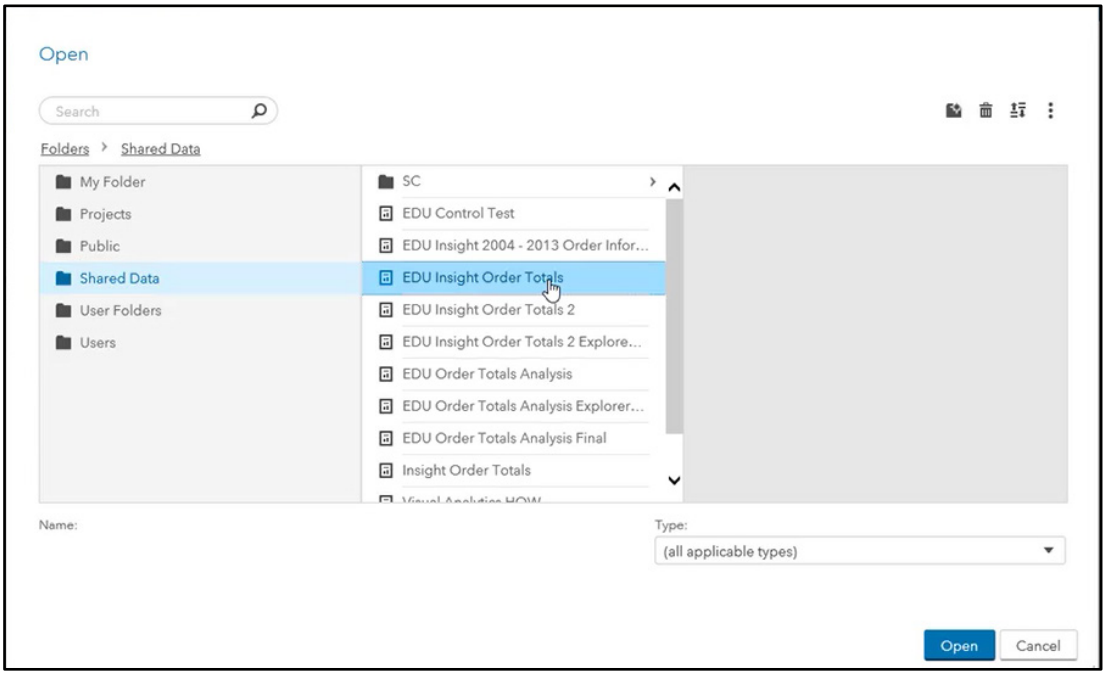
Edit a Tile

Next, let's add a bookmark to the Favorites tile. Select the Options menu and then Edit, as shown in Figure 1.6.

Figure 1.6: Favorites Tile Options Menu

In the Edit Favorites window, click the + sign, which is also called the Add icon. This pulls up the Open window, as shown in Figure 1.7. In the Open window, select the report that you want to add as a favorite, and click Open.

Figure 1.7: Choose Report Window



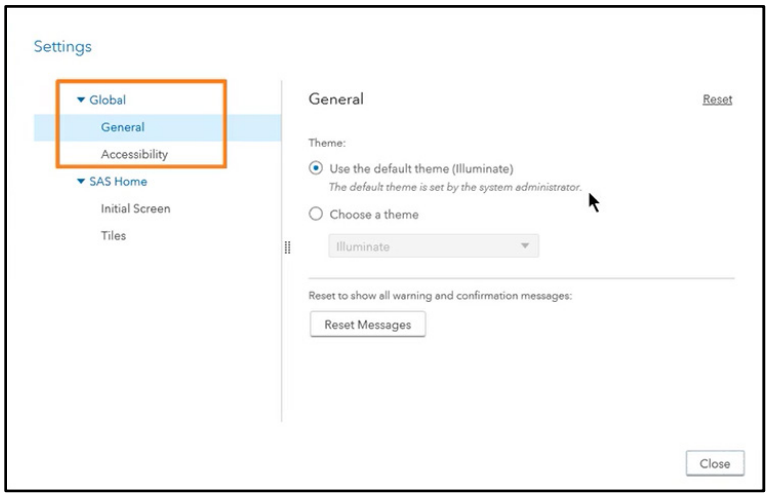
Back in the Edit Favorites window, select Save, and the bookmark for the report will now be added to the favorites tile, as shown in Figure 1.8

Figure 1.8: Updated Favorites Tile

Application Bar Settings

Finally, let's modify some of the settings for the Home page. Click your sign-in name in the upper right-hand corner of the application bar at the top. Here, you can select Settings or Sign Out. We will select Settings. This opens the Settings window as shown in Figure 1.9.

Figure 1.9: Settings



Global Settings

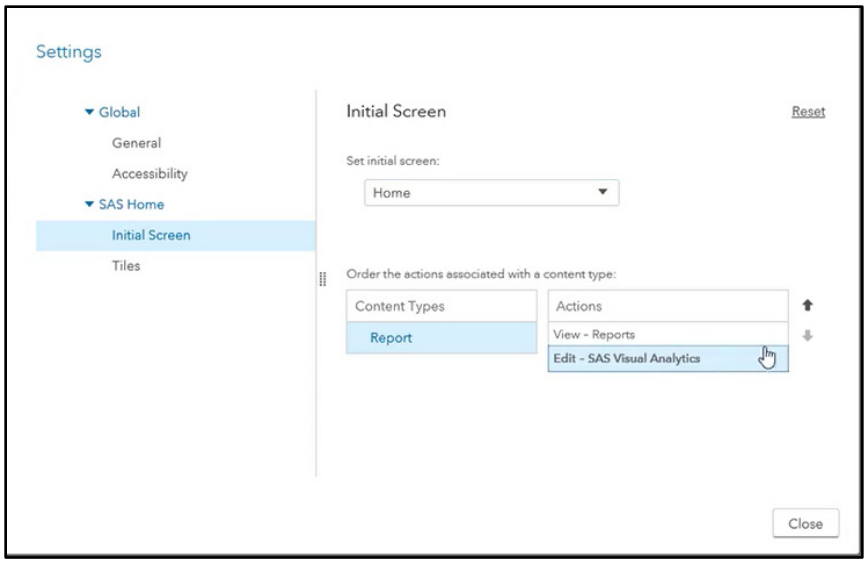
In the Settings window, there are a couple of global and accessibility settings. We will not go into these in detail here. Instead, we will move on to the next section on the left, which is the SAS Home page.

SAS Home Page

For the SAS Home page, there are settings for the initial screen and tiles. The Initial screen setting enables you to select the screen that opens when you sign in. The default is the Home page, but depending on your role, you might want to open another application by default by selecting that application from the drop-down menu.

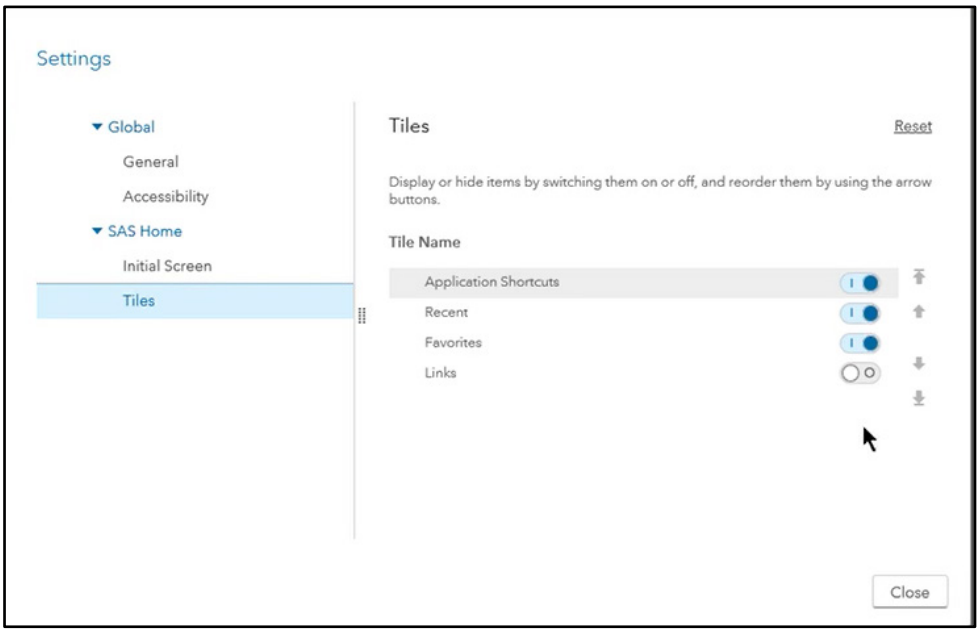
You can also change the order of actions for a report. The default action is listed at the top, and here the default is to open a report in the viewer. An analyst who can edit or view a report might prefer to set the default action to open in the editor. To make this change, you select the action that you want as the default and use the arrow keys to the right to move it to the top, as shown in Figure 1.10.

Figure 1.10: Initial Screen Settings



Now let's look at the settings for Tiles. See Figure 1.11.

Figure 1.11: Tile Settings

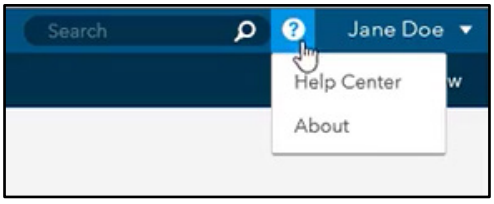


You can choose to hide or show tiles by clicking on the toggle button to the right. If one of the tiles has been hidden, you will be able to see that it is turned off here. You can also control the arrangement of tiles on the Home page by selecting a tile and using the up or down arrows on the right to specify the order. When you are finished in the Settings window, click Close to exit.

Help

To see documentation for the Home page and other features in SAS Visual Analytics, click the Help menu button in the upper right-hand corner of the top banner, and then select Help Center to open the SAS documentation in a new window, as shown in Figure 1.12.

Figure 1.12: Help Center

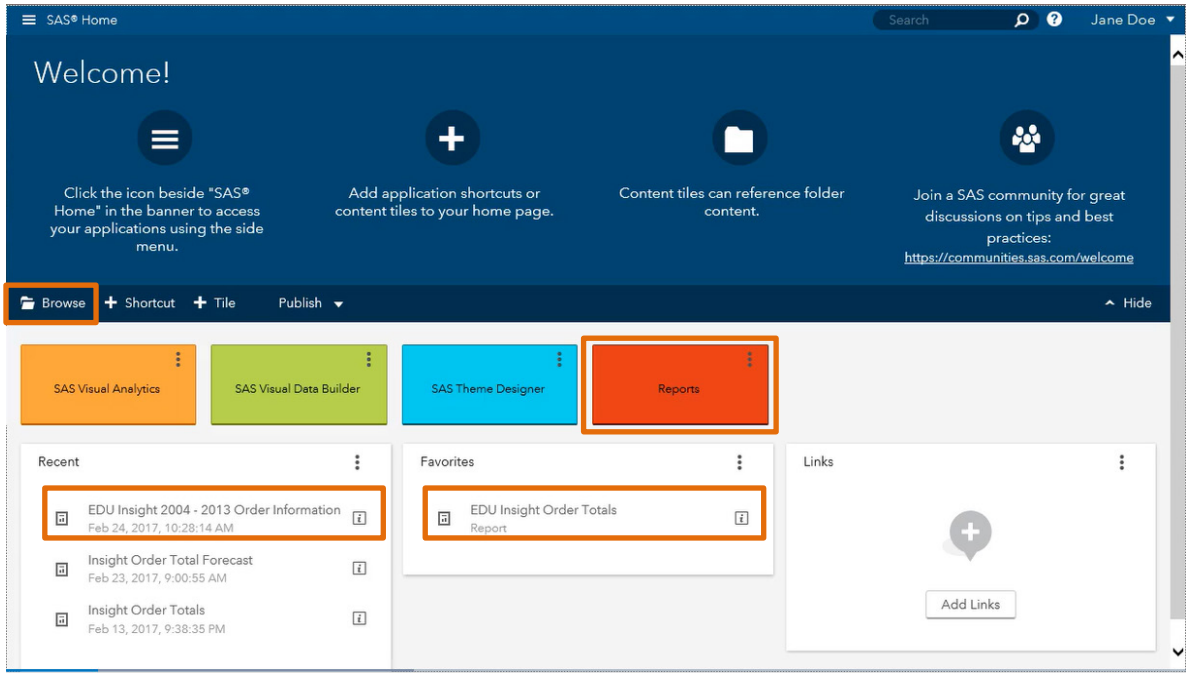


Using SAS Report Viewer

In this section, we will learn how to open, view, and use actions in an existing SAS Visual Analytics report in the SAS Report Viewer. For this example, the report that we want to open is called "EDU Insight 2004–2013 Order Information Report." The sample data used to create this report is available from the SAS Visual Analytics 8.1 on SAS Viya 3.2 Documentation page.

From your Home page, there are multiple ways to open a report in the Report Viewer. Each of the following methods is highlighted in a box in Figure 1.13.

Figure 1.13: Methods to Open Report Viewer from Home Page

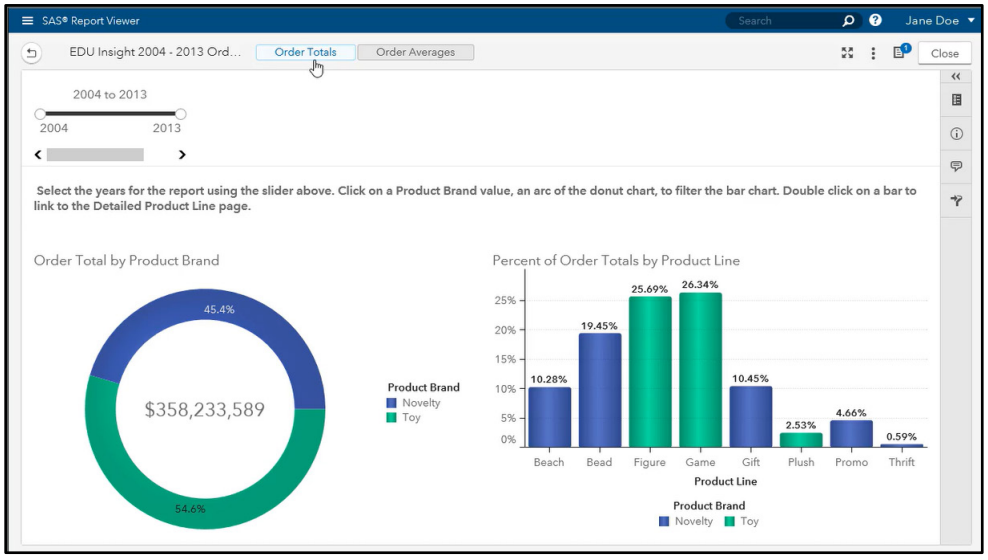


The methods shown in Figure 1.13 are:

- **Browse**. Select Browse from the menu bar and navigate to the SAS folder that contains the report and open it from there.
- **Recent Tile**. If you recently viewed the report, you can use the link in the Recent tile to open it.
- **Favorites Tile**. If you added a bookmark to the report in your Favorites tile, you can use the bookmark to open it.
- **Report Application**. You can also open the Report Viewer application directly by selecting the shortcut button included on the Home Page. To find a report, you can enter a term in the filter field to limit the reports displayed in the list. Or you can browse through all folders or browse recently viewed reports.

Once you open the report with any of these methods, you will see the report in SAS Report Viewer, as shown in Figure 1.14.

Figure 1.14: SAS Report Viewer



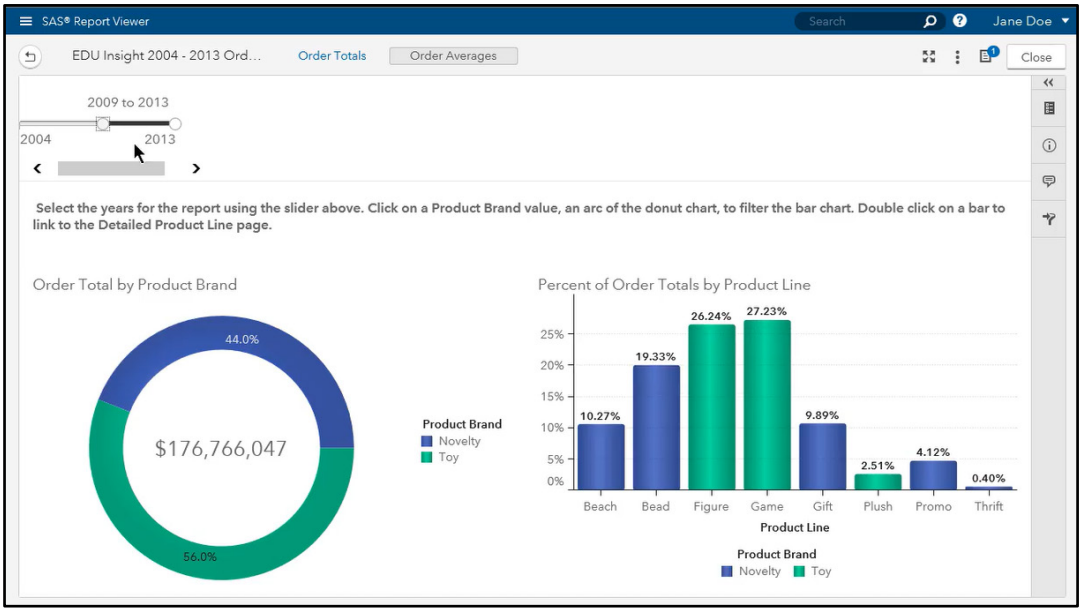
Notice that the report has two pages: Order Totals and Order Averages. This report also has a filter that you can use to choose the number of years of data to include in the report. Below the filter are instructions about the actions that are included in the report, as well as a donut chart and a bar chart.

Slider

Looking at the donut chart in Figure 1.14, we can see that over the 10-year period from 2004 to 2013, the Toy brand accounts for 54.6% of the total orders. What will happen if we want to look at only the last 5 years of data? Will that percentage change? Move the year slider to change the beginning year to 2009 and see what happens.

When the year slider is moved, a new query is passed to the in-memory table and results are updated as shown in Figure 1.15. The year slider is a report-level prompt, so it is applied to all pages in the report.

Figure 1.15: Slider Feature

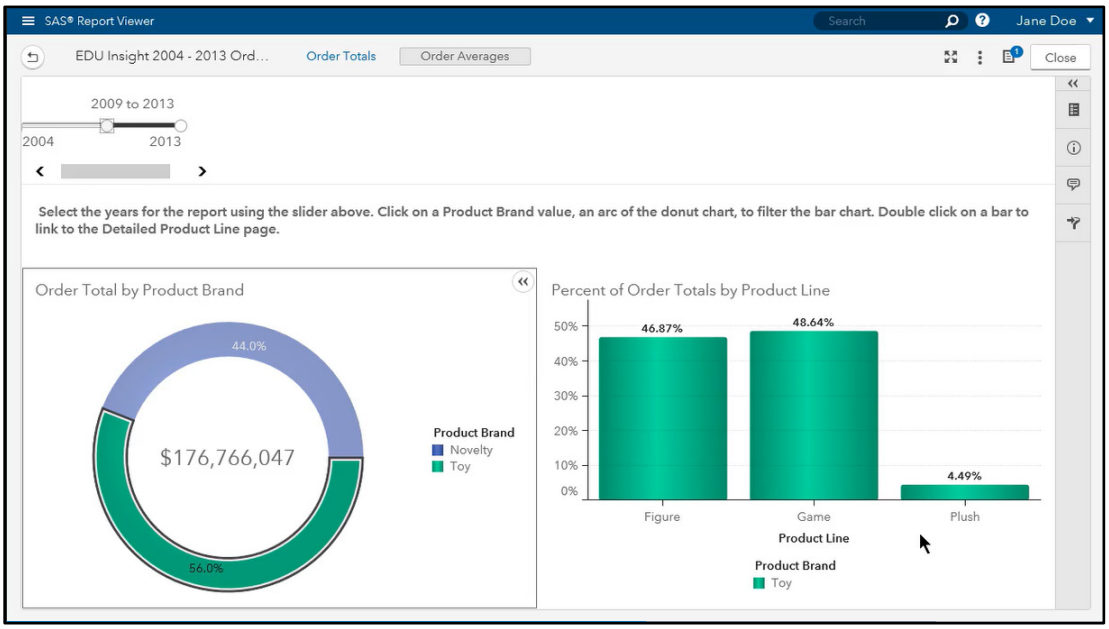


As shown in Figure 1.15, we can see that over the past 5 years, the toy brand accounts for 56% of total orders.

Filter

The instructions in the report indicate that there is a filter action set up between the donut chart and the bar chart. When we click the Toy arc, the bar chart displays only the product line values from the Toy brand, as shown in Figure 1.16.

Figure 1.16: Filter Action

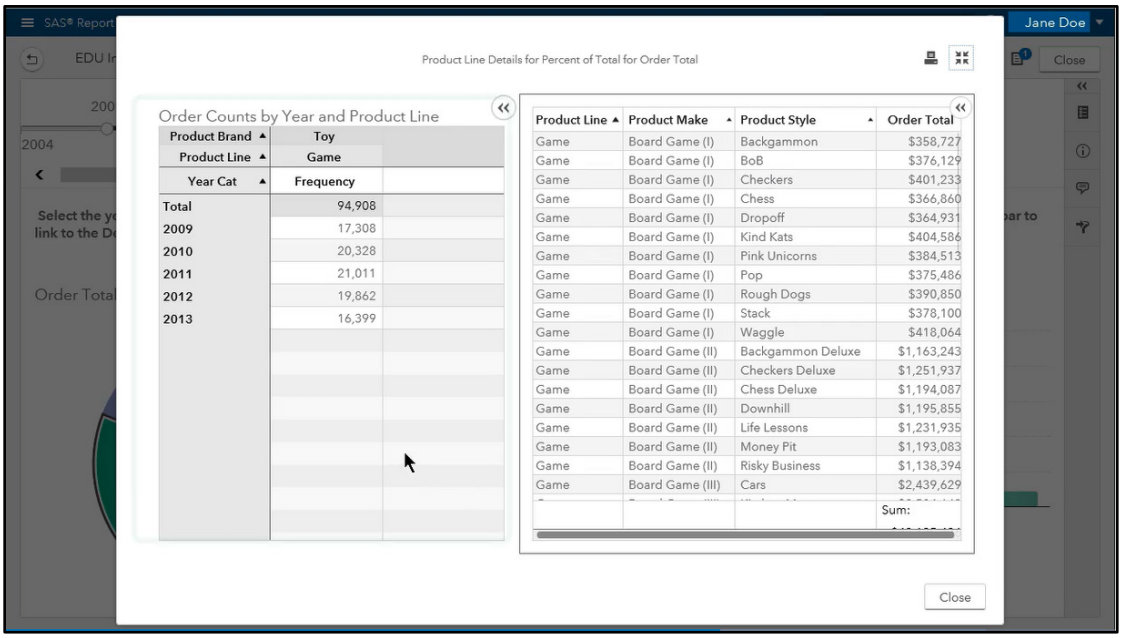


Details

The instructions indicate a further action. By double-clicking a bar in the bar chart, we can see a link to the Detailed Product Line Page.

If you double-click the Game bar, a window opens with two tables, as shown in Figure 1.17. To increase the size of the window, click the maximize button in the upper right-hand corner.

Figure 1.17: Details Window

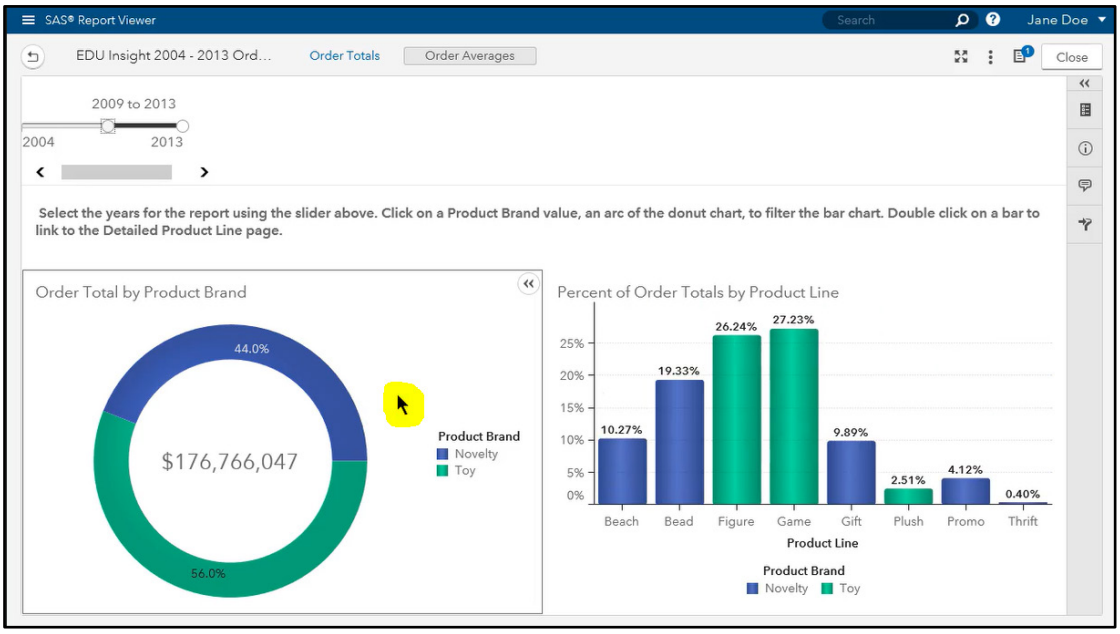


The table on the left is a crosstab that displays the order counts by year for the Game product line. The table on the right is a detailed listing of the Game product line data. Close the window by selecting the Close button in the lower right-hand corner.

Clear Filters

Back on the Order Totals page, you can clear the filters on the page by clicking in the white space by the donut chart, as shown in Figure 1.18 where the pointer is highlighted. Notice that this only resets the filters; it does not reset the slider.

Figure 1.18: Clear Filters

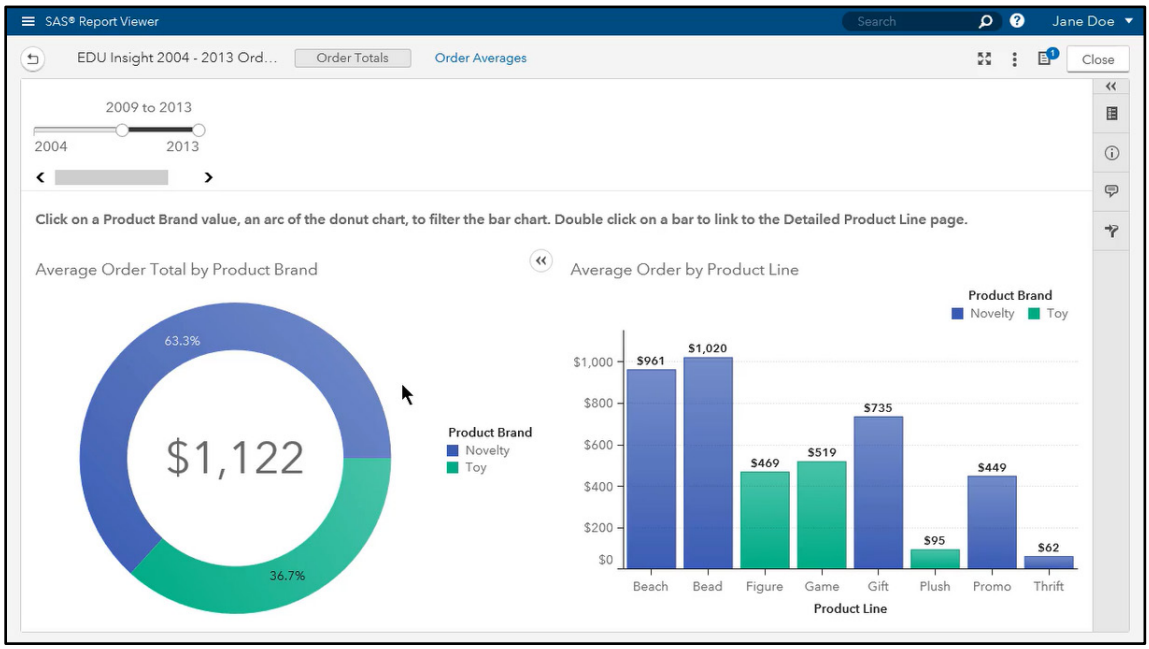


Information

Now, let's look at the second page of the report by clicking Order Averages at the top of the screen to move to the next page. As you will recall, the year slider is a report-level prompt, so it is applied to all pages in the report.

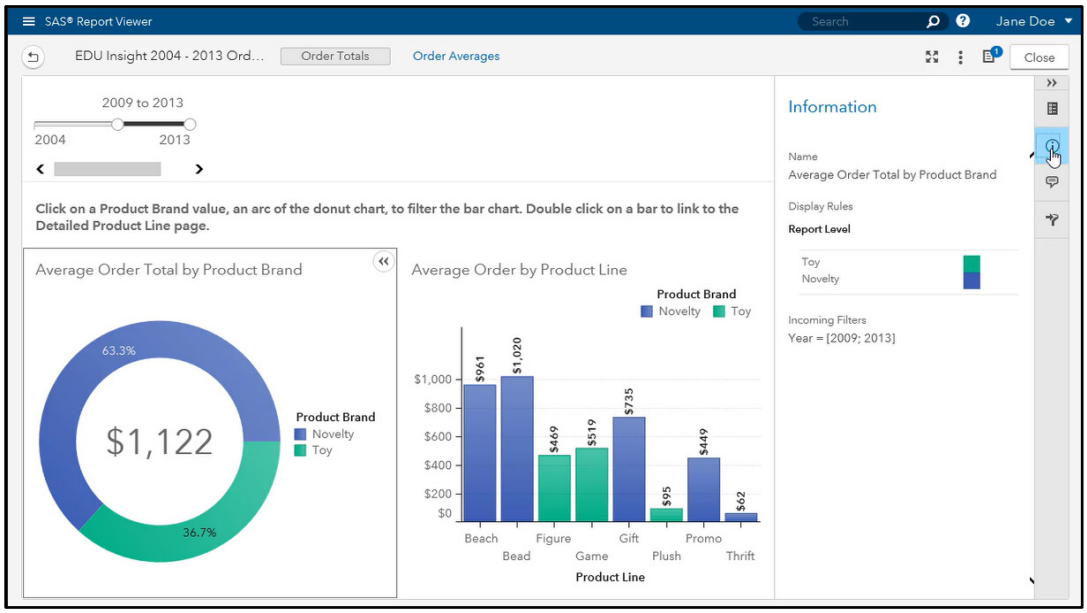
On the previous Order Totals page, we saw that the Toy brand accounted for a higher percentage of the order total, or revenue. But on this page, we can see that the average per order is higher for the Novelty brand, as shown in the donut chart in Figure 1.19.

Figure 1.19: Order Averages Page



If you want to see more information about an item in the report, you can click the information icon in the right-hand pane, as shown in Figure 1.20. This opens the information pane.

Figure 1.20: Information Pane



Here, we see that the donut chart is filtered by years. To close this panel, click the information icon again.

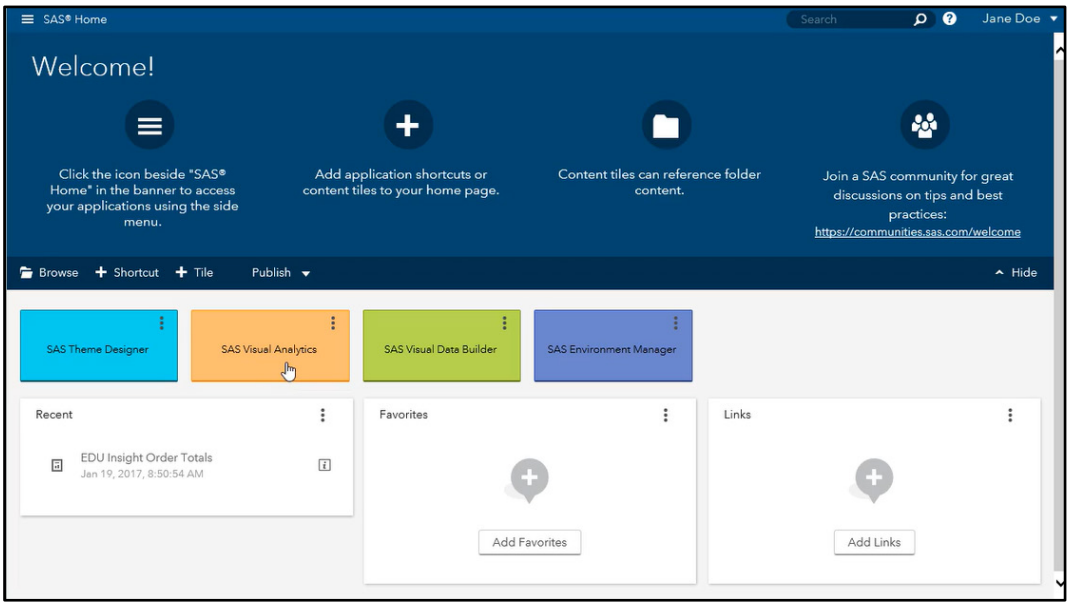
Close Report

To close a report, click Close in the upper right-hand corner of the report. To return to your Home page if you are not already there, click the side menu in the upper left-hand corner, and select Home.

Creating a Basic Report

Now that you have customized your Home page and viewed some of the features of a report, let's learn how to create a report in SAS Visual Analytics. When you are signed in to your Home page, click the SAS Visual Analytics tile shortcut, which is shown as an orange tile in Figure 1.21.

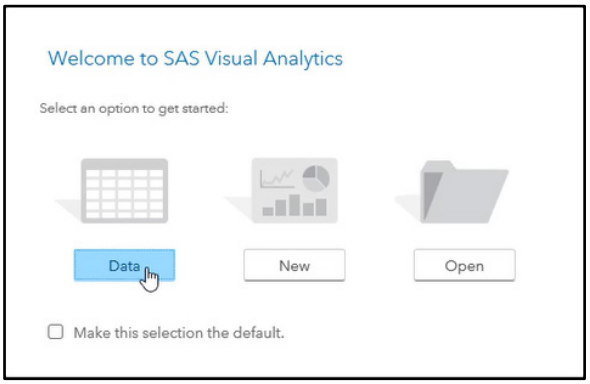
Figure 1.21: Home Page



On the welcome screen that is shown in Figure 1.22, you will see the following three options for starting your report:

- **Data**. Select a data source.
- **New**. Start a new report without adding data.
- **Open**. Browse for an existing report.

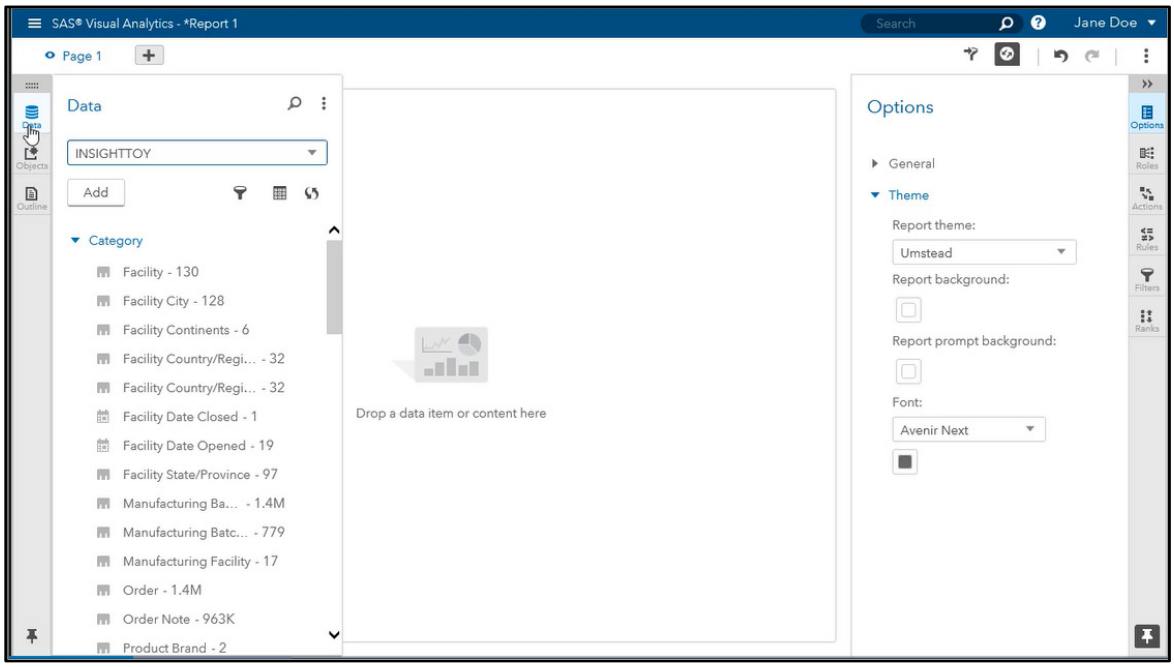
Figure 1.22: SAS Visual Analytics Welcome Screen



In this example, we will start our report by selecting a data source, so we click the Data button. The Open Data Source window appears. We will create a report from a data set called INSIGHTTOY. In the search bar, search for the name of the data set. Select INSIGHTTOY from the Open Data Source window. Then click OK.

The data source has been added to your report and you are in design mode as shown in Figure 1.23.

Figure 1.23: Design Mode



As you design your report, you will be using three main areas. The far-left pane has three buttons that toggle open the Data, Objects, and Outline panels. Currently, in Figure 1.23, the Data panel is open. In the middle of the screen is the canvas where you will design your report. The far-right pane has buttons for opening the Options, Roles, Actions, Rules, Filters, and Ranks panels. Currently, in Figure 1.23, the Options panel is open.

Generally, when you are designing a report, any changes that you make in the left-hand pane will apply to *all* parts of the report. Any changes that you make in the right-hand pane will apply only to *individual* report objects.

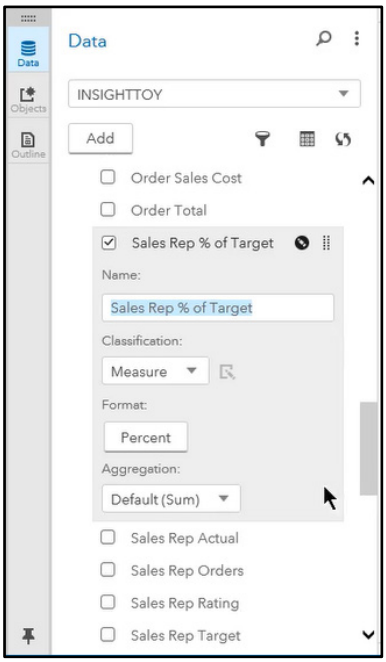
Data Panel

Let's look more closely at the Data panel on the far-left side of the screen. In Figure 1.23, you see a list of data items in the INSIGHTTOY table. The data items are arranged alphabetically within the classification type. There are two classification types: Category and Measure. The Category classification consists of character variables and date variables. The Measure classification consists of the numeric variables that we can analyze.

In the list of data items, you can highlight a data item by hovering over the name to see more information. For Category data items, you can see the number of distinct values and the format that is applied. For Measure data items, you can see the format and the aggregation that are applied.

As part of the preparation for creating a report that has both summary and detail information for orders in the product lines sold by Insight Toy, we will edit a data item. Hover over the Measure item "Sales Rep & of Target." Then select the Edit properties for this data item, which appears as a black circle with a pencil as shown in Figure 1.24.

Figure 1.24: Edit Data Properties

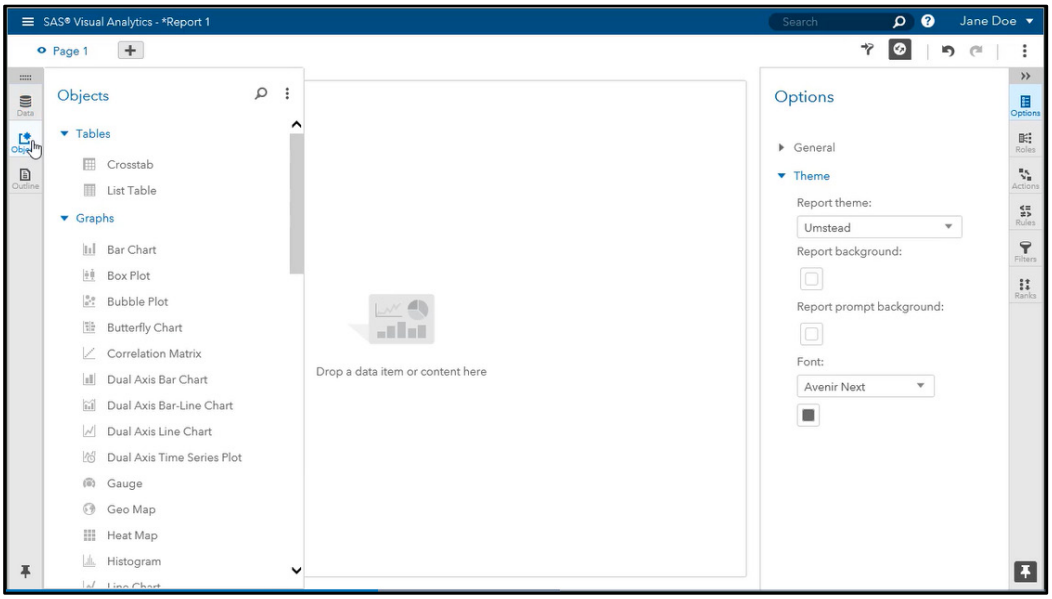


Let's look at how the properties are defined for this data item. The default aggregation is Sum. For this data item, Average is more meaningful. So select Sum and use the drop-down menu to change the aggregation to Average.

Objects Panel

Now that you have seen how to view and edit data items in the Data panel, we can begin to add report objects by using the Object panel. Select Objects in the far-left pane to view the Objects panel as shown in Figure 1.25.

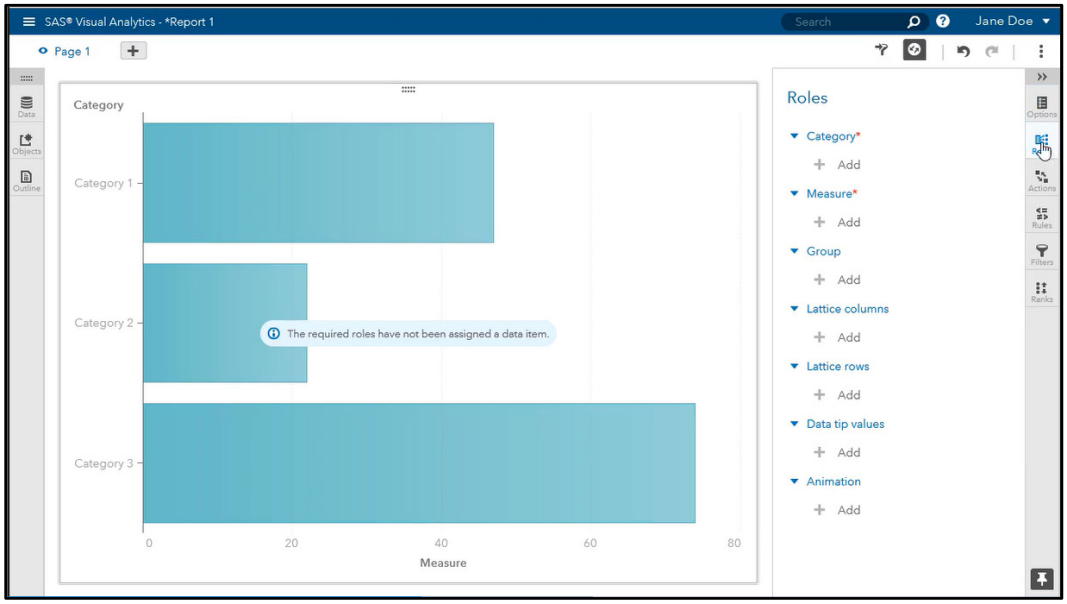
Figure 1.25: Object Panel



Add an Object (Roles)

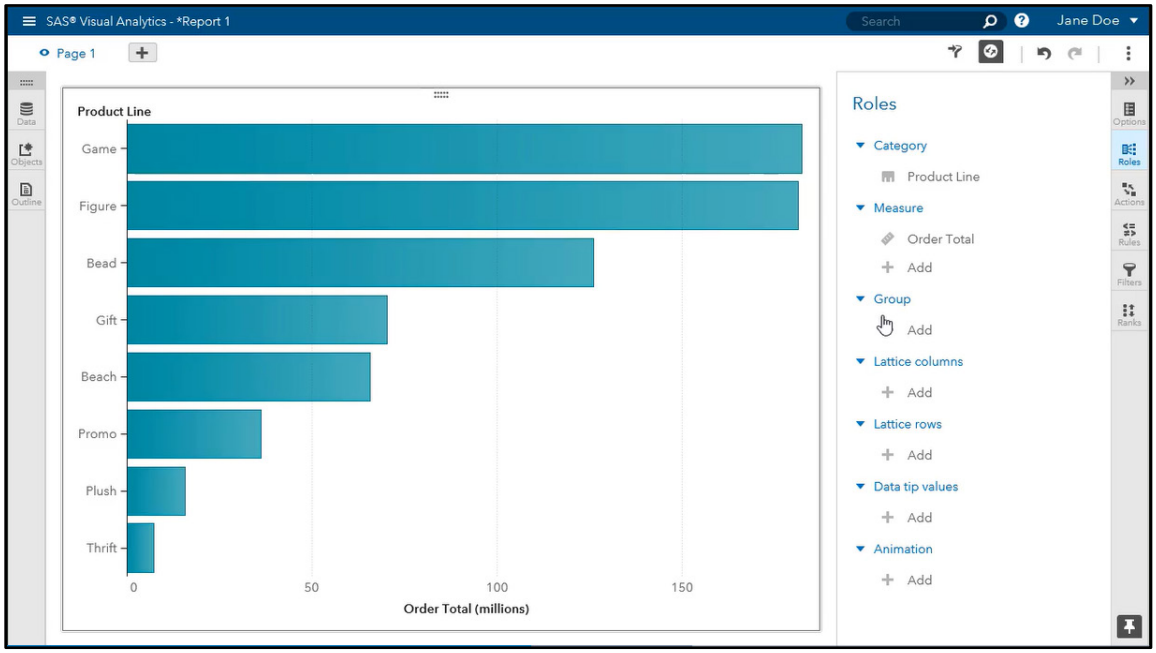
To add a bar chart to the report, select the Bar Chart object from the Object panel, and drag it over to the canvas in the middle of the screen. A placeholder bar chart appears in the canvas area, as shown in Figure 1.26.

Figure 1.26: Placeholder Chart



Now we can associate data items with roles in the bar chart. In the far-right pane, select Roles. You can tell from the red asterisks which roles are required for the bar chart. Under Category, select Add. Then choose Product Line from the window that appears. The bar chart updates automatically, with Frequency used in the Measure role by default. However, we want to compare order totals for the product lines, so we will change the Measure. Click Frequency, and in the Replace Data Item window, select Order Total instead. Our bar chart now looks like the chart shown in Figure 1.27.

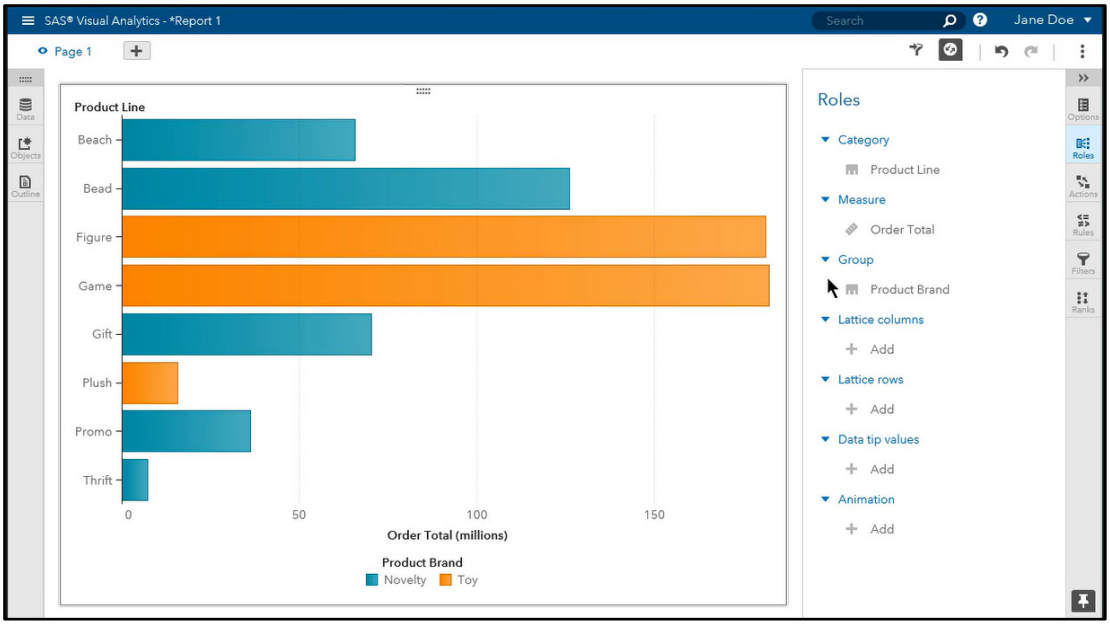
Figure 1.27: Product Line Bar Chart



We can see that Game and Figure are the two top product lines. What if you want to see this chart grouped by Brand? For the Group Role in the right-hand pane, select Add, and then choose Product Brand.

Now we can see that the top two product lines, Game and Figure, are in the Toy brand, as shown in Figure 1.28.

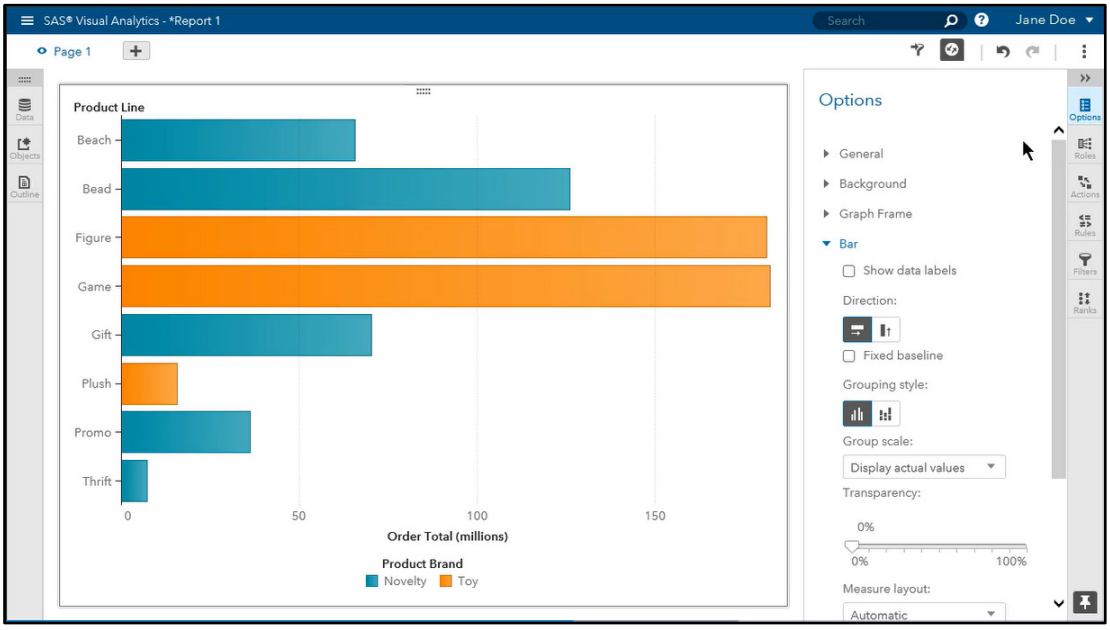
Figure 1.28: Product Line Bar Chart Grouped by Brand



Modify Properties (Options)

Next, let's modify some of the properties for our bar chart. In the far-right pane, click Options. This will cause the Options panel to appear on the right, as shown in Figure 1.29.

Figure 1.29: Options Panel



To add a title, click General and enter the title of your chart in the Title field. For this example, enter "Order Totals by Product Line" in the title field, and then press Enter.

In the Bar section, select the check box for Show data labels. Increase the text size to 10 and set the Bold attribute. In the Bar section, you can also change the direction of the bars to vertical instead of horizontal.

In the Y Axis options section, clear the check box for Axis label.

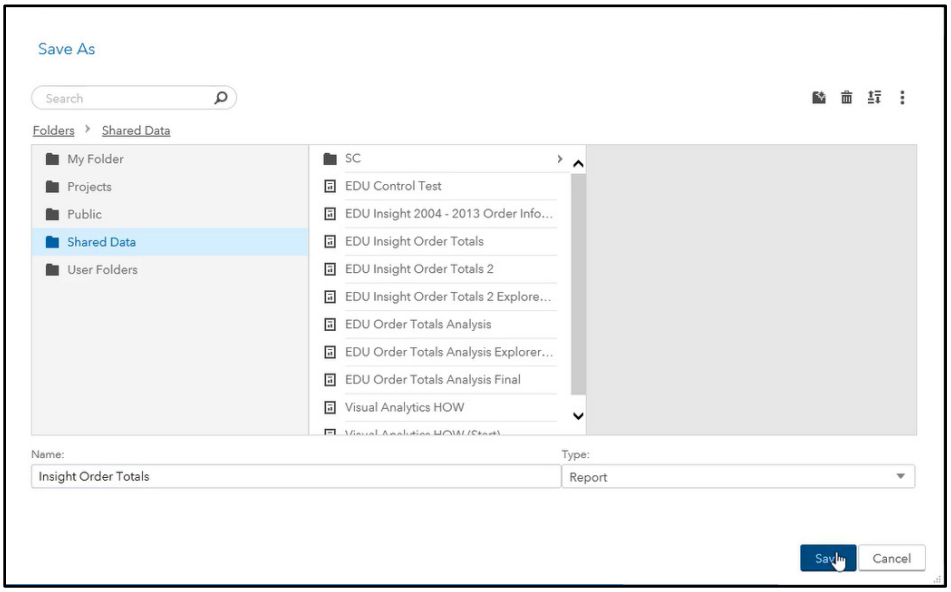
Finally, in the Legends section, change the placement to the upper right-hand position by selecting the radio button that corresponds to that location.

After modifying all those options, your bar chart will look like the chart shown in Figure 1.30.

Figure 1.30: Order Totals by Product Line

It looks pretty good, right? To save the report, click the More Options in the upper right-hand corner of the screen, which appears as three vertical dots. Select Save As. In the Save As window in Figure 1.31, you can choose the location to which you want to save your report. In this example, select the Shared Data folder, and in the Name field, enter "Insight Order Totals." Select Save.

Figure 1.31: Save As Window



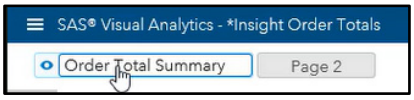
Congratulations! You have just created your first report.

Add a Page

In this section, you will learn how to add a new page to the report that you created in the previous section. Click the + sign, which is the Add a Page button to the right of Page 1 at the top right-hand corner of the screen. A blank page appears, and you see a Page 2 heading at the top of the screen.

Now that you have multiple pages, it might be helpful to give each page a meaningful name. Double-click the button for Page 1 and rename it "Order Total Summary" as shown in Figure 1.32. Double-click Page 2, and name it "Order Total Details."

Figure 1.32: Rename Page

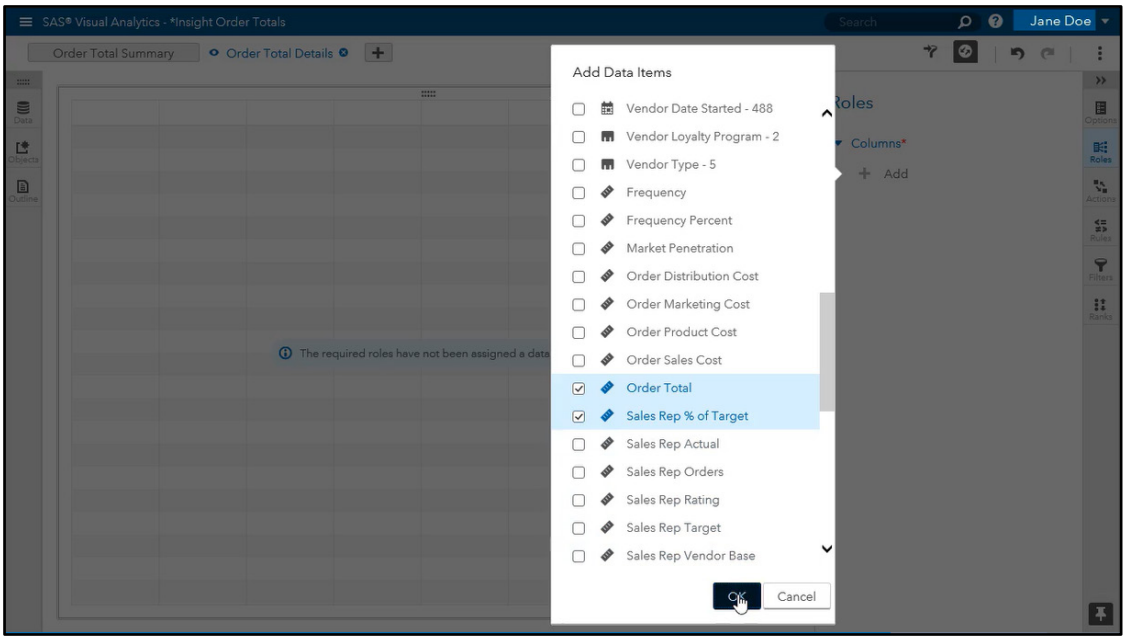


On this new page, you can follow the previous steps that we used to create a report by examining the data and adding objects. In this example, we will not edit any of the data, but we will add a list table that shows detailed data for orders.

In the left-hand pane, select Objects. Drag the List Table object to the canvas. You will not see a placeholder chart.

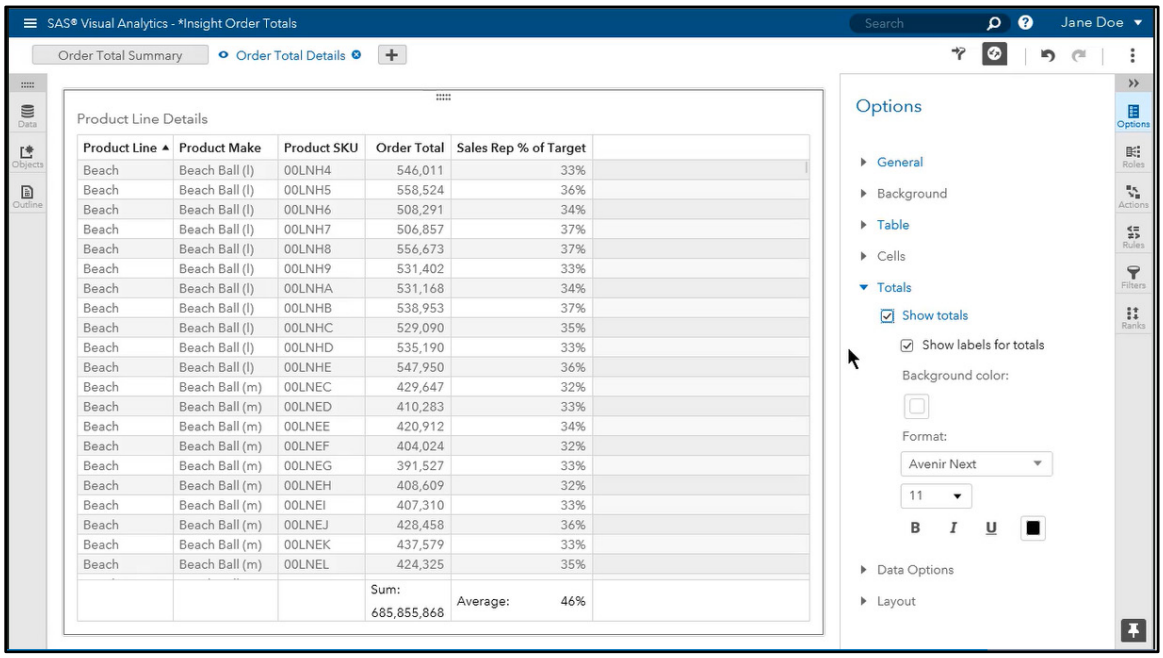
In the far-right pane, select Roles to select columns for the list table. Click Add, and then select the five following data items: Product Line, Product Make, Product SKU, Order Total, and Sales Rep % of Total, as shown in Figure 1.33. Then click OK.

Figure 1.33: Add Data Items for List Table



Add a Title to this table by clicking Options in the far-right pane. In the Title field, enter the title "Product Line Details." Also, expand the Totals section, and select Show Totals, as shown in Figure 1.34.

Figure 1.34: Options Panel for List Table



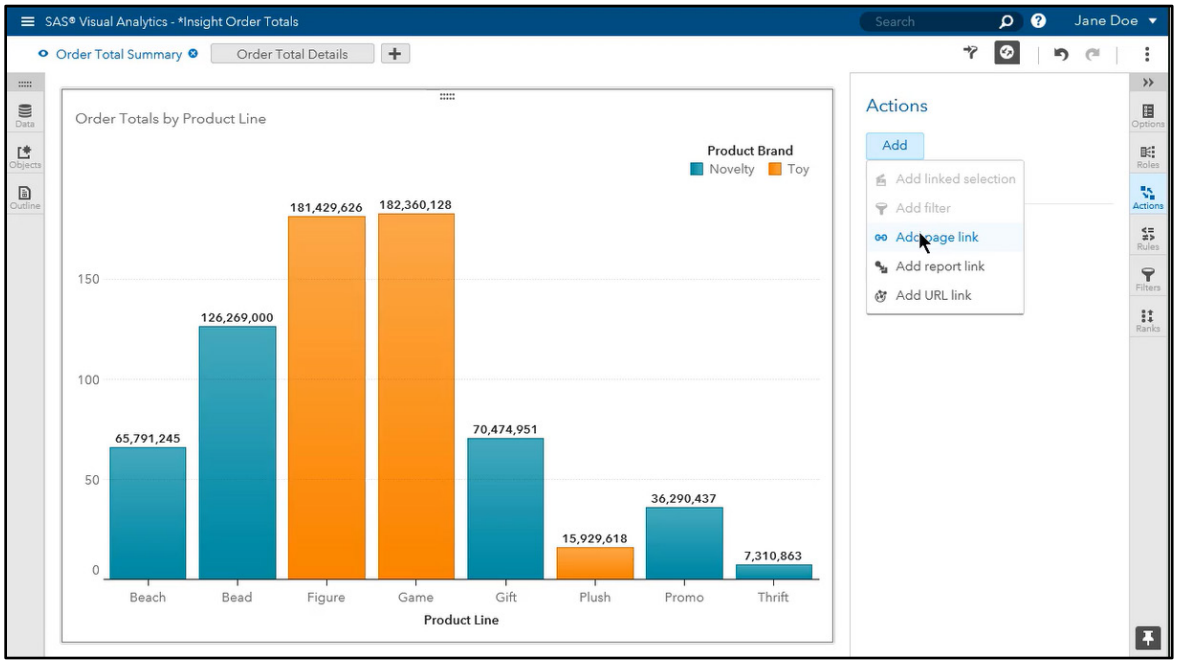
There are many rows in this table, so let's connect the two pages in our report with the filter action.

Connecting Pages

Click "Order Total Summary" to return to the first page of the report. Remember in the previous section that this is what we renamed "Page 1." Now, let's select the bar chart to make it active.

In the far-right pane, select Actions, and click the Add button. From the menu, select Add page link as shown in Figure 1.35.

Figure 1.35: Add Page Link



In the Add Page Link Actions window that appears, you can select which pages you want to link to. Here, you simply click OK because there is only one other page in this report. The link to the other page now shows in the Actions pane.

Let's test the action by double-clicking the bar for the Beach product line. This page link action automatically flips the page to display the Order Total Details page and filters the list table object to display rows for only the Beach product line.

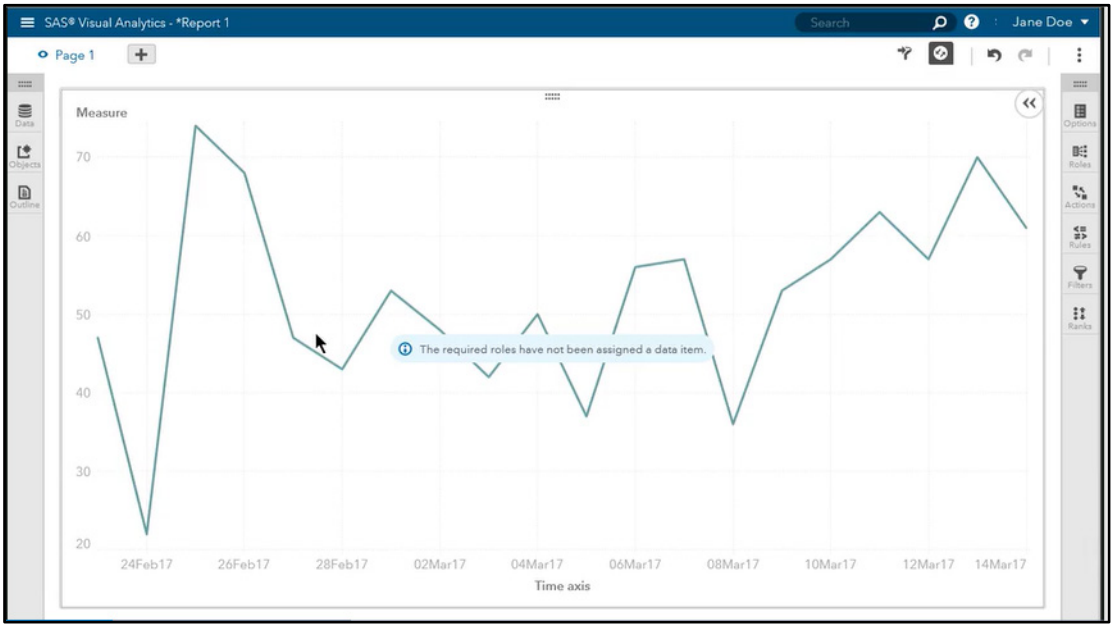
Save this report by clicking the More Options icon, which appears as three vertical dots, in the upper right-hand corner of the screen. Select Save.

Creating an Advanced Report with a Forecast

Let's use SAS Visual Analytics to further explore the sample data used in this chapter, the INSIGHTTOY data, and learn how to use some more advanced options to create new columns. In this section, we will focus on the key metric, Order Total, and then use that data to generate a forecast of the next 6 months.

After opening SAS Visual Analytics, select the INSIGHTTOY data source from the data panel. In the Objects panel, select the Time Series plot object, and drag it to the canvas. A placeholder time series plot appears in the window as shown in Figure 1.36.

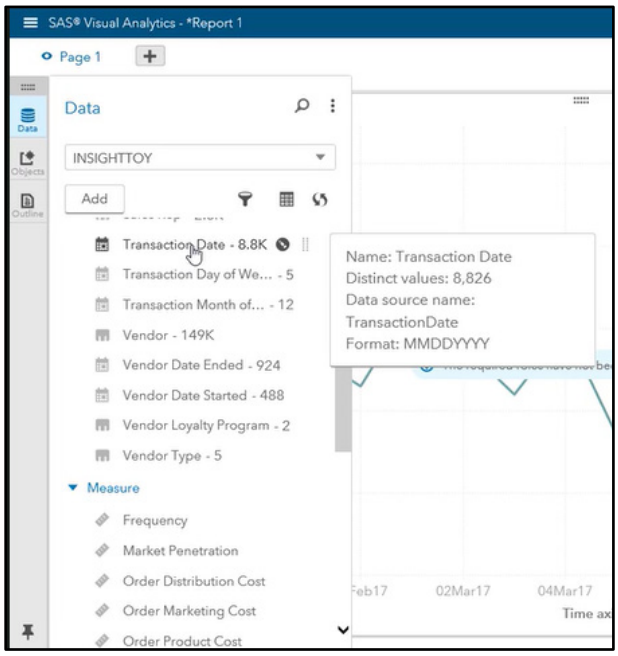
Figure 1.36: Placeholder Time Series Plot



Modifying Time Values

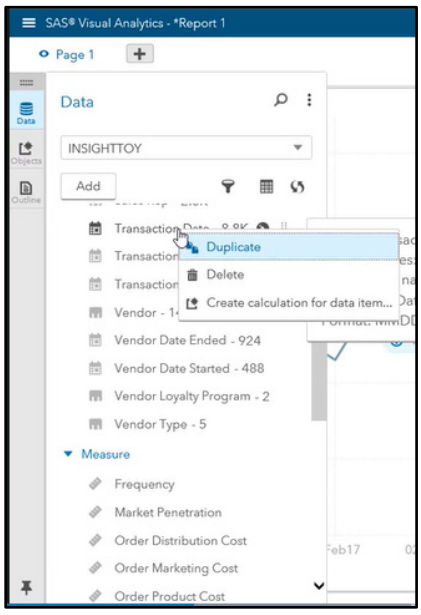
A time series plot requires a date formatted column on the Time, or X, axis. When we return to the data panel and look at the INSIGHTTOY data, we see that there are three data values, as shown in Figure 1.37: Transaction Date, Transaction Day of Week, and Transaction Month of Year.

Figure 1.37: Data Panel



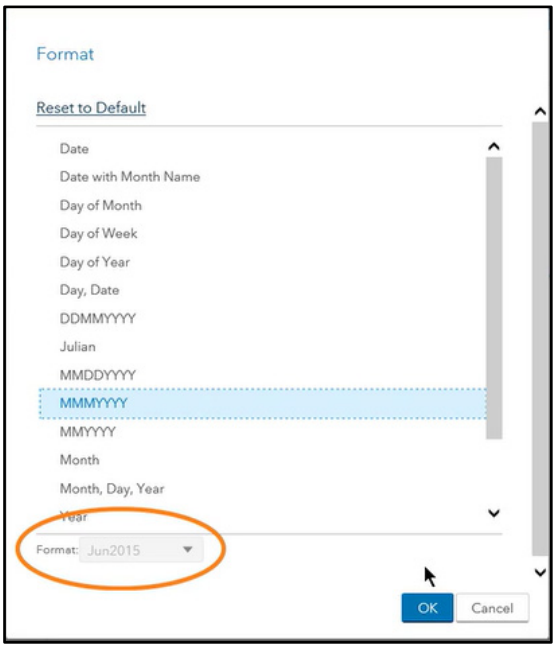
For this example, we want to use a date with a different format than the ones already available in the data. To do this, we will duplicate a column and select the format that we need. To duplicate a column, right-click the name of the data value, and click Duplicate, as shown in Figure 1.38.

Figure 1.38: Duplicate a Column



Notice that a new column has been created with the name Transaction Date (1). Click the Edit Properties button for the new column, Transaction Date (1). Change the name to “Transaction Month.” Then, click the Format field, which currently shows MMDDYYYY. A new format window appears, as shown in Figure 1.39.

Figure 1.39: Format Window



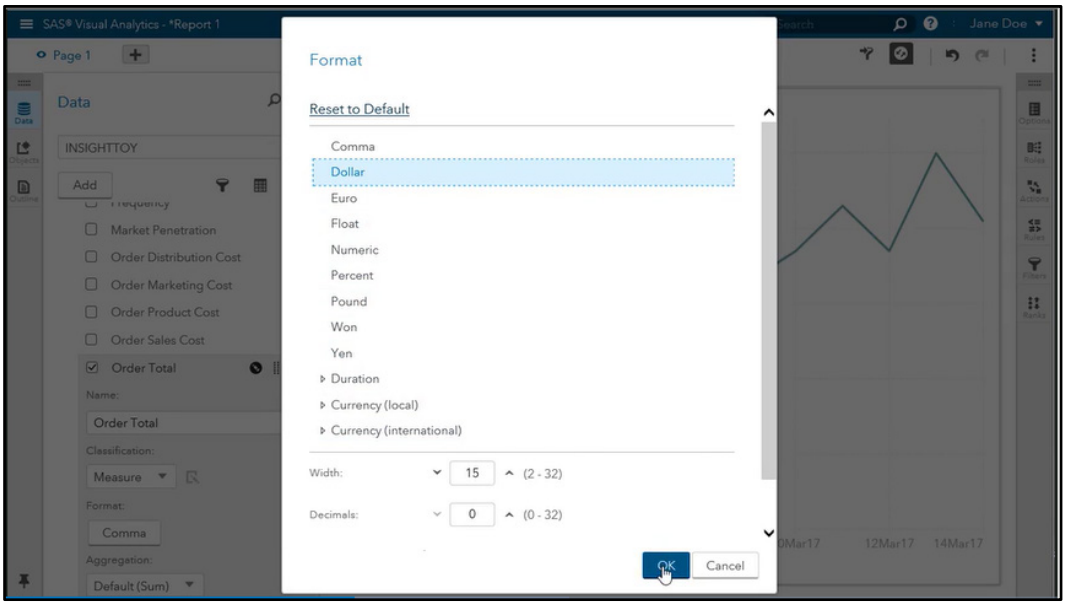
In the Format window, change the format to MMMYYYY. In the format display circled in orange in Figure 1.39, you can see how your new format will be displayed as a separate data point for each month by year. This is the format that we need to generate a monthly forecast. Click OK.

Click Edit Properties again on the newly named Transaction Month column to close the edit panel.

Modifying Measures

For the next part of this example, we will modify the Y-axis properties for the Order Total column to reflect that the data is in US dollars. Click the Edit Properties button on the Order Total column in the Data panel. Click Format, and then change the format from Comma to Dollar as shown in Figure 1.40.

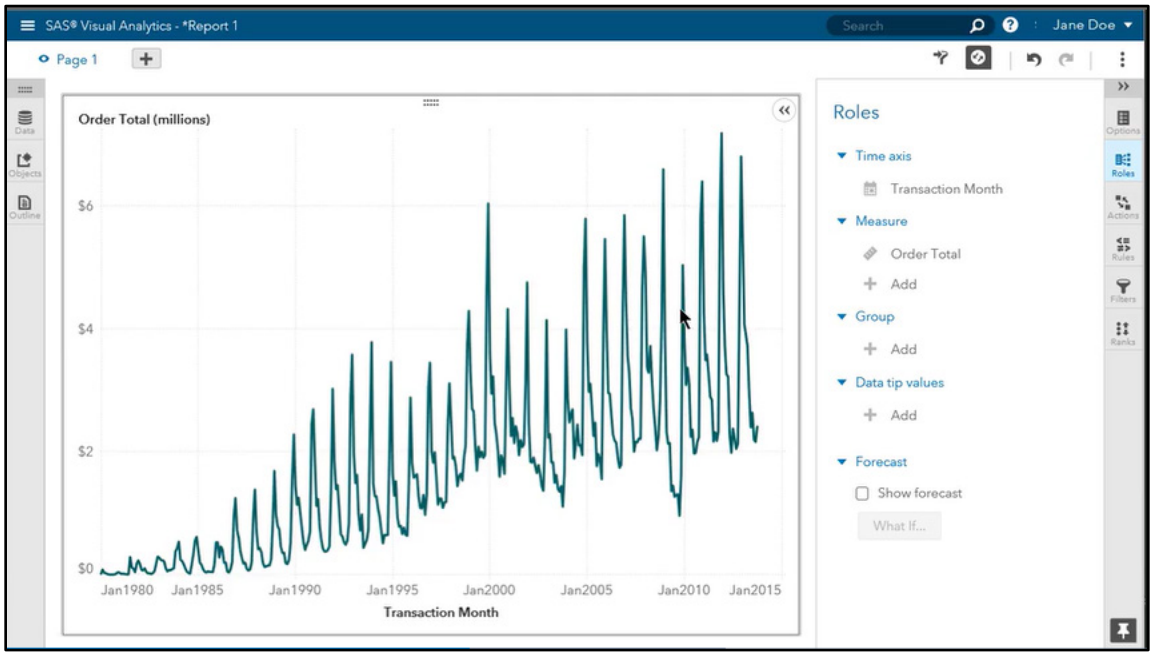
Figure 1.40: Format



Assigning Items to the Report

Now we can select the data items for the roles in the time series plot. Select Roles in the right-hand pane. For the Time Axis role, select Add, and then select the new column that we created earlier, Transaction Month. For the measure, select Frequency, and then replace it with Order Total, which we modified earlier to be formatted in US dollars. After you make these selections, your time series plot looks like the one shown in Figure 1.41.

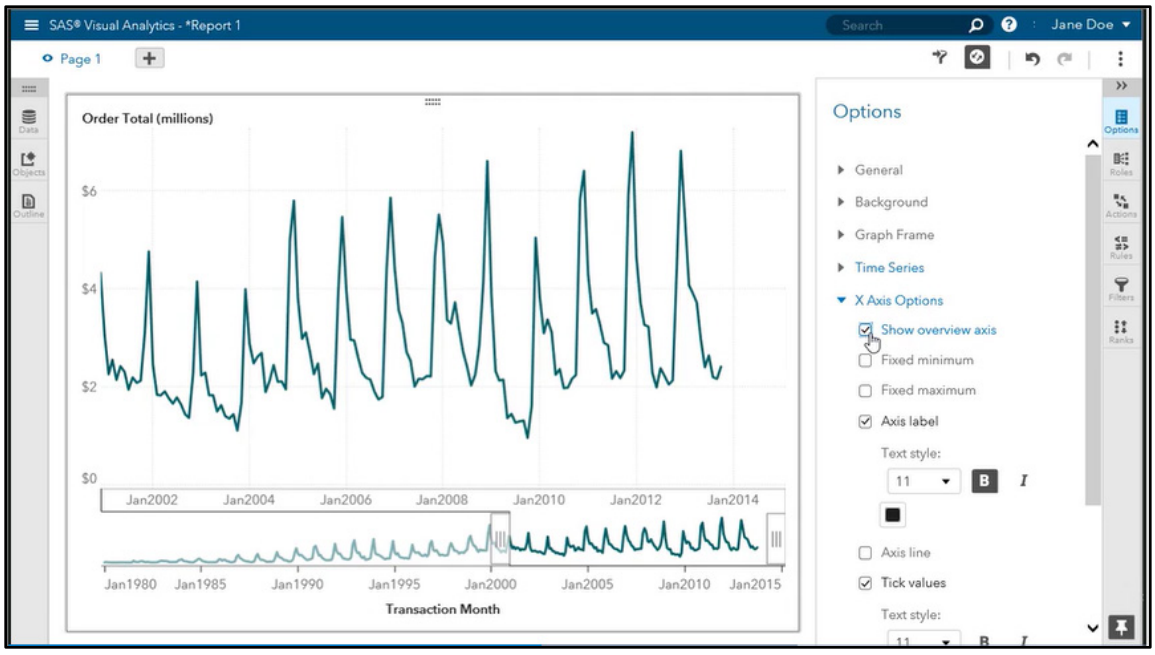
Figure 1.41: Time Series Plot



You might notice that the time series plot shows more than 30 years of data. You can add a filter to limit the number of years, but you can also change the display.

Click Options in the right-hand pane, expand X Axis Options, and select Show overview axis, as shown in Figure 1.42. With the overview axis, you can control the length of time that is currently displayed in the plot and focus on a briefer time period.

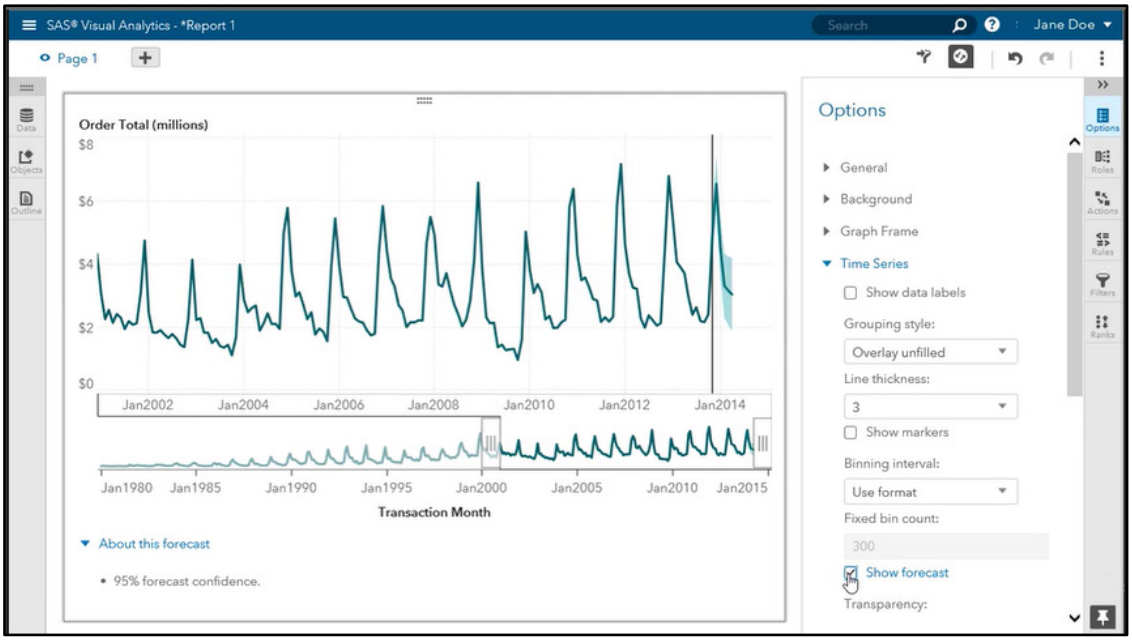
Figure 1.42: Add Show Overview Axis Option



Adding a Forecast

Now let's add a forecast to our time series plot. In the Options panel on the right, expand Time Series, and click the Show forecast check box, as shown in Figure 1.43.

Figure 1.43: Add Show Forecast Option



To focus on the past 5 years of available data, adjust the left slider on the overview axis to show data only from January 2009 to 2013. Because we have added a forecast, a Forecast section is added to the Options panel, as shown in Figure 1.44. Here, you can modify the confidence interval or the number of periods in the forecast horizon. Currently, the number is set at 6, which is a 6-month horizon.

Figure 1.44: Options Panel Forecast Section

Suppose that this forecast is the one that you want to share with your manager. Save it by selecting the More Options in the upper right-hand corner of the screen (three vertical dots), and then select Save As. In the Save As window, select the Shared Data folder, and name the report "Insight Order Total Forecast."

Explore Mode

To understand the details of this forecast, let's view this plot in Explore mode. Close all the panels on the right and left side of the screen. Click the Object Options menu (<<) in the upper right-hand area of the window, and then select Explore as shown in Figure 1.45.

Figure 1.45: Open Explore Mode

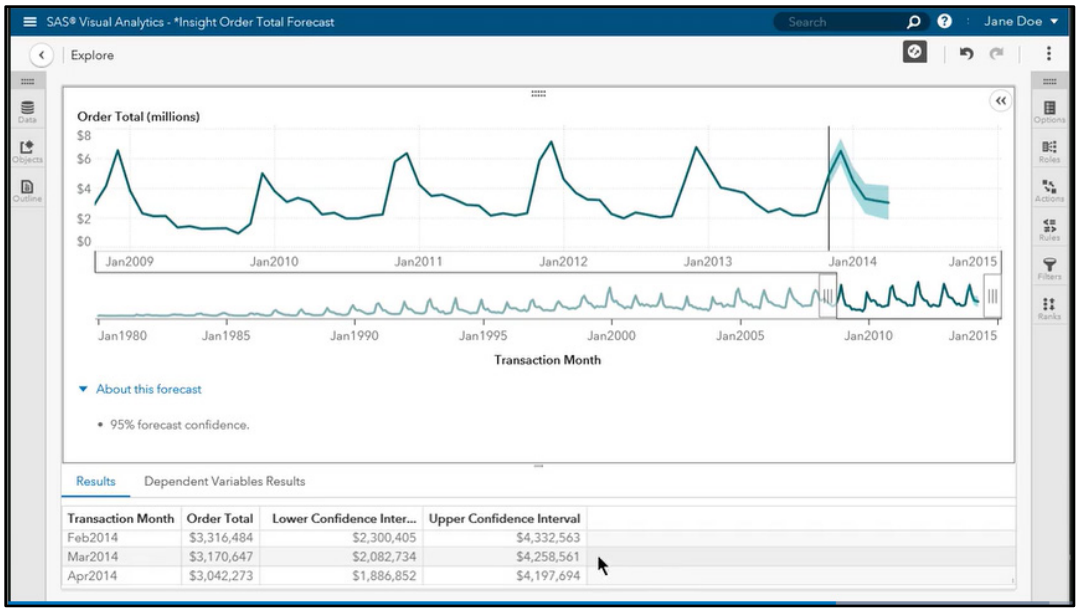
In this mode, the plot has two tabs below it:

- Results
- Dependent Variable Results

Results Tab

In the Results tab, as shown in Figure 1.46, you can scroll to the bottom, where you can see the forecasted values for November 2013 through April 2014, as well as the lower and upper confidence intervals for these values.

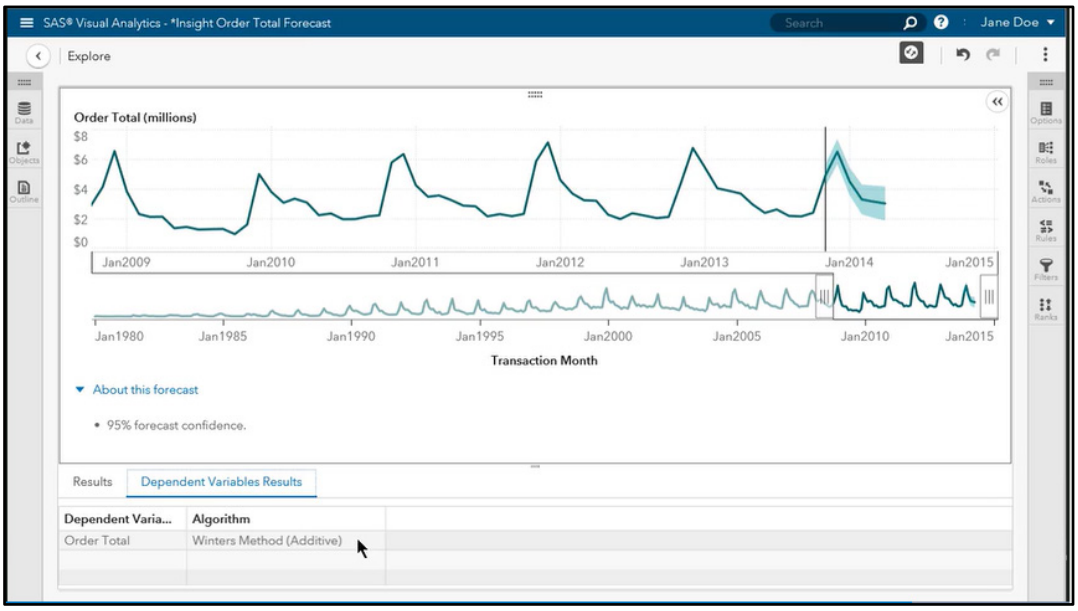
Figure 1.46: Results Tab



Dependent Variables Results Tab

In the Dependent Variables Results tab in Figure 1.47, you can see that the best algorithm selected to forecast this data was the Winters Method (Additive). Now you have the answers to the questions that you might receive from your manager about this analysis.

Figure 1.47: Depend Variables Results Tab

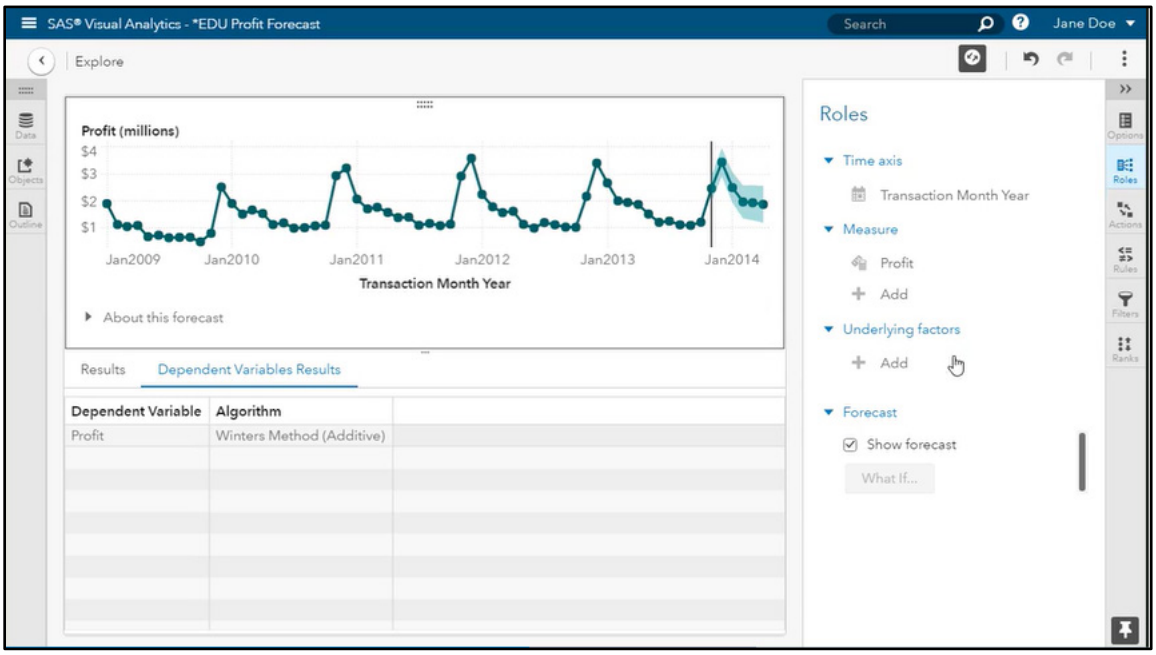


If you save the report while you are working in Explore Mode, then the next time you open the report in Visual Analytics, it will open in Explore Mode again. To avoid this, exit Explore Mode by clicking the Return to Report arrow button in the upper left-hand corner, next to the word *Explore*. Now you can save the report by clicking More Options and Save.

Filters

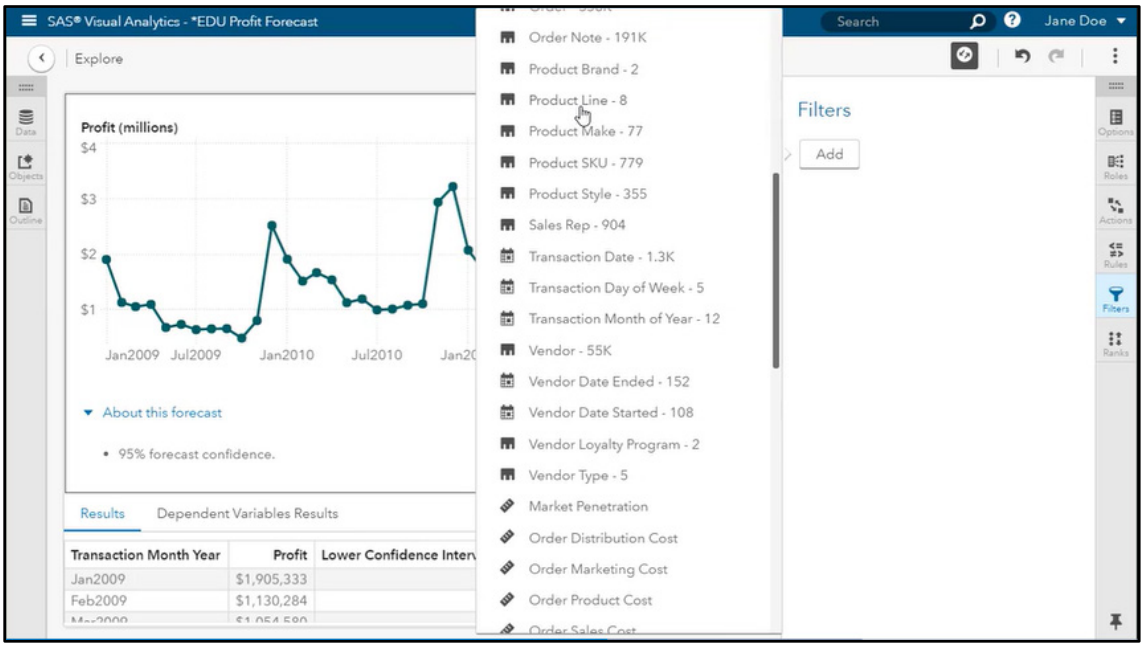
The plot in Figure 1.47 shows us profit for all products, but maybe you want to see how some of the individual product lines performed. One way to do so is by adding Product Line to the Group role. In the right-hand pane, click Roles. Notice in Figure 1.48 that the Group role is not available. This is because the time series plot contains a forecast.

Figure 1.48: Roles Panel



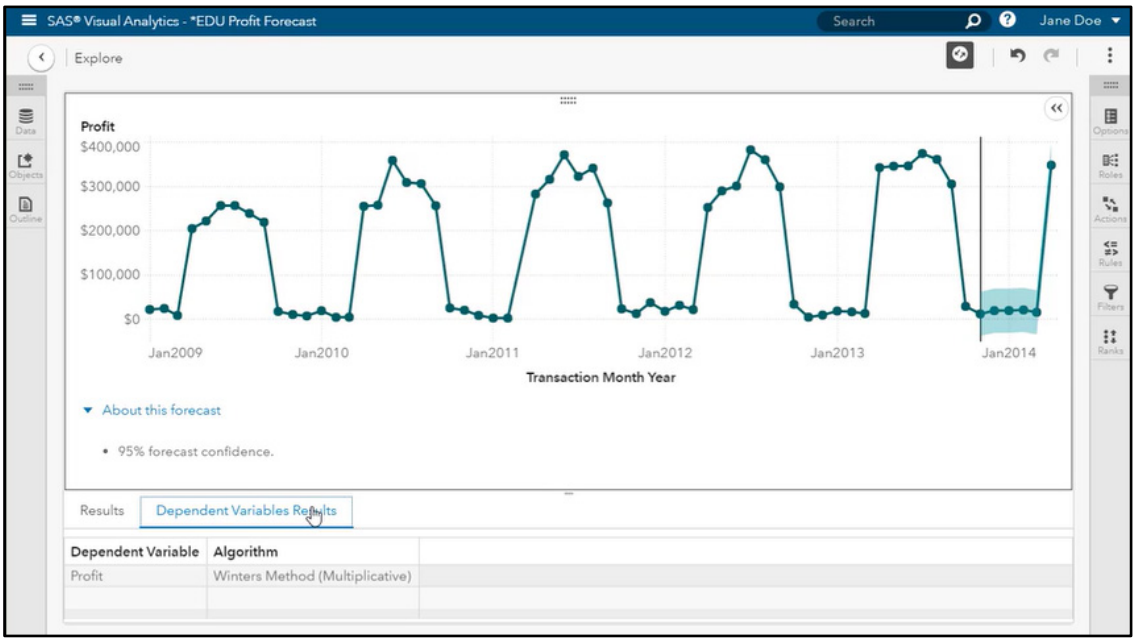
Instead of adding a Group role, you can add a filter for Product line. Click Filters, select Add, and select Product Line, as shown in Figure 1.49.

Figure 1.49: Add a Filter



Now, we can look at the time series plot for individual product lines. Start by clearing the check box for Select All and then selecting Beach. The plot shown in Figure 1.50 is quite different from the plot for all product lines, but it is not surprising that profit for Beach products is higher in the summer than in the winter. You might also notice that the algorithm for the forecast is different. This set of data was best forecasted with use of the Winters Method multiplicative algorithm.

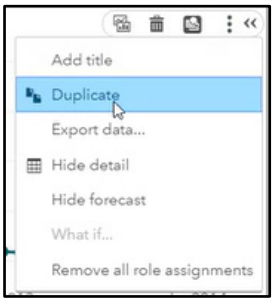
Figure 1.50: Beach Product Line Time Series Plot with Forecast



Create New Graph Objects

What if we want to compare the Beach product line to the other product lines? An easy way to do so is to duplicate this time series plot. Open the Object Menu, and then select More Options and then Duplicate, as shown in Figure 1.51.

Figure 1.51: Duplicate Option

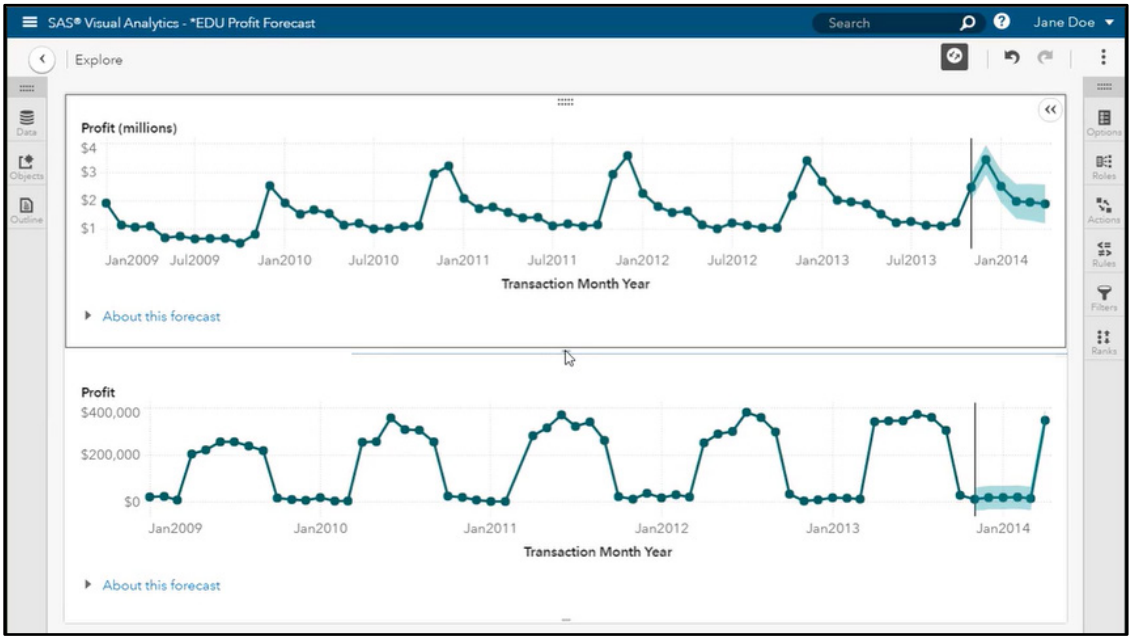


A copy of the forecast is added below the original with all the same attributes, including the filter that we just added, as shown in Figure 1.52.

Figure 1.52: Duplicated Forecast

To change the filter on the original chart, follow the same steps to add a filter as before. Click Filters in the rightmost pane, and then click the Select All check box to include all product lines. You can now see that the top forecast in Figure 1.53 includes all product lines, while the bottom includes the Beach product line.

Figure 1.53: Duplicated Forecast with Filters

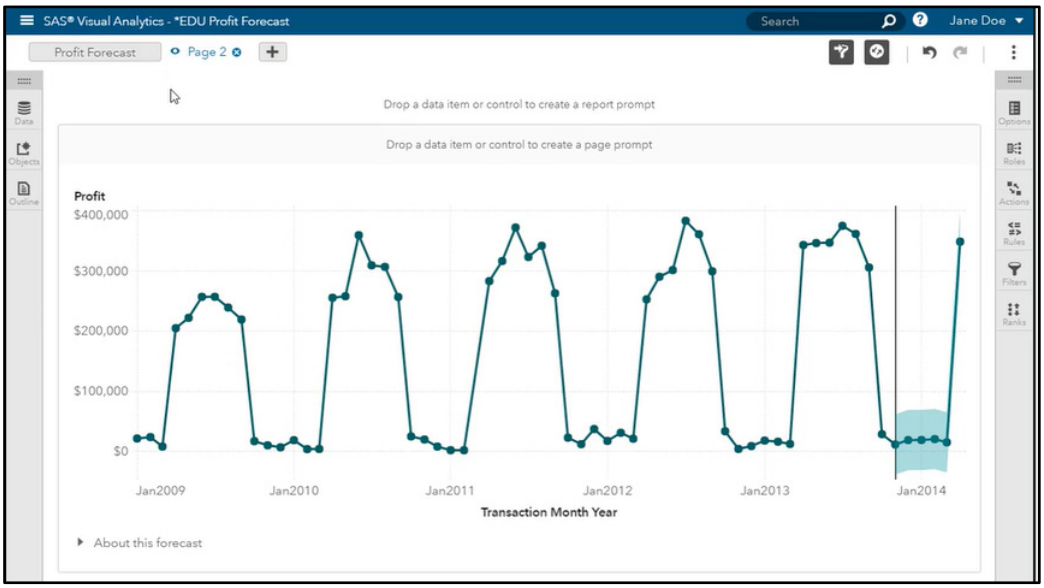


We have already noticed that the profit for the Beach product line is highest when other product lines are not as profitable. If we look at the scale of the Y axis on each chart, we can see that the profit from the beach product line is relatively small compared to overall profit. Perhaps the company might consider adding some additional seasonal products to increase profit during the less profitable times of the year.

We could spend a lot of time comparing different product lines by changing the filter combinations in the two charts. But let's exit Explore mode and return to the report by clicking the arrow in the upper left.

When we exit Explore mode, we are back to the original report layout. The chart that we added is on Page 2 as shown in Figure 1.54, and the page is hidden.

Figure 1.54: Modified Report



When a page is hidden, it is not visible to anyone who views the report in the Report Viewer. We could leave the page hidden, or unhide it to include it in the report. Use the eye icon on the left side of the page button to hide or unhide. We can also delete the page by using the x icon on the right side of the page button.

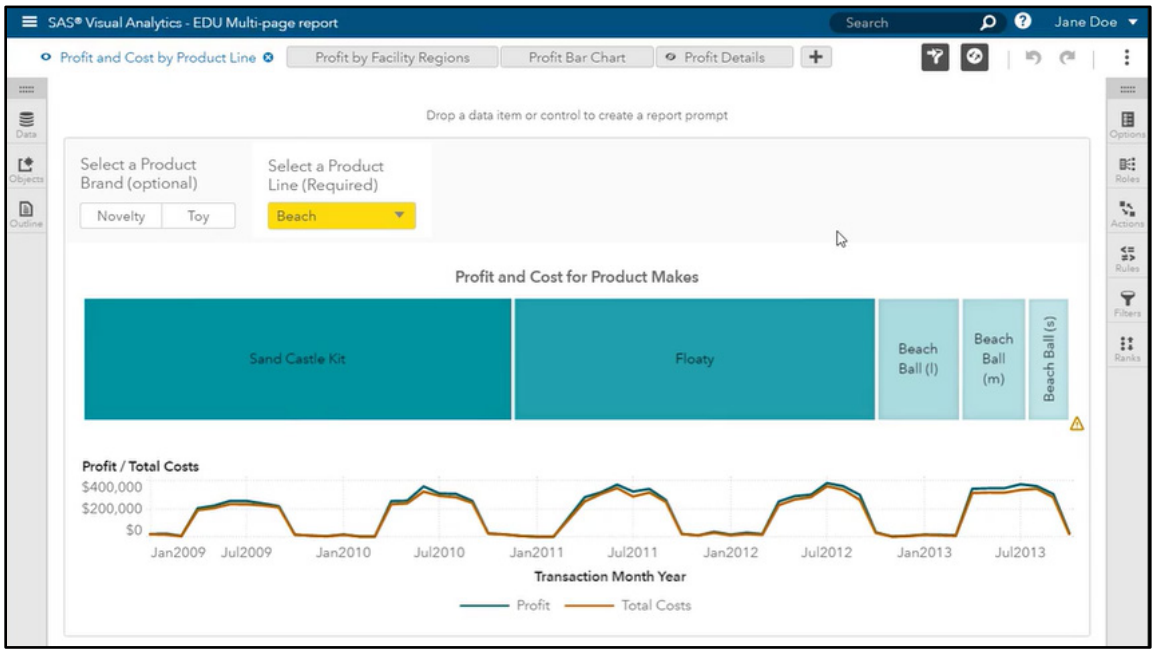
In this section you learned how to add a forecast to a time series plot and view information about the forecast. We have seen how to use Explore mode to examine additional information about charts and create new objects. When we create new objects in Explore mode, the original layout of the page is preserved, and the new content is displayed on a new hidden page in the report.

Modifying Reports

In this section, we will be looking at an example report that has four pages with various analyses of profit for a toy company.

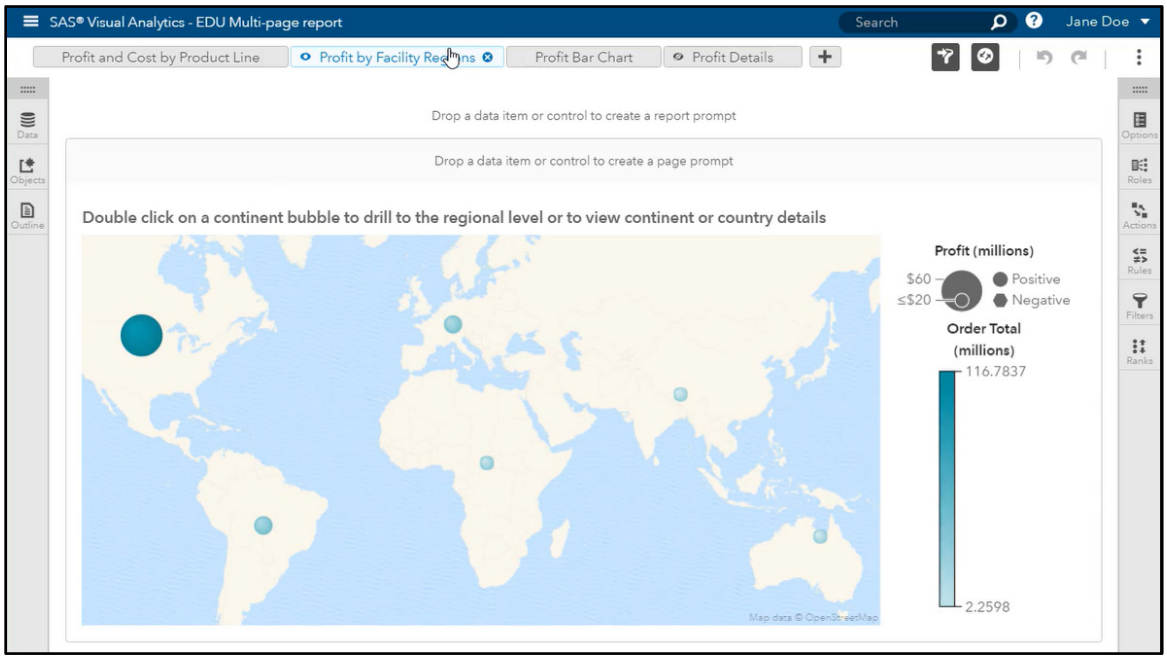
The Profit and Cost by Product Line page shown in Figure 1.55 has two prompts that filter the data in the treemap and time series plot on the page.

Figure 1.55: Profit and Cost by Product Line



The next page, the Profit by Facility Regions page shown in Figure 1.56, has a geographical map with bubbles that show Profit and Order Total for each continent.

Figure 1.56: Profit by Facility Regions

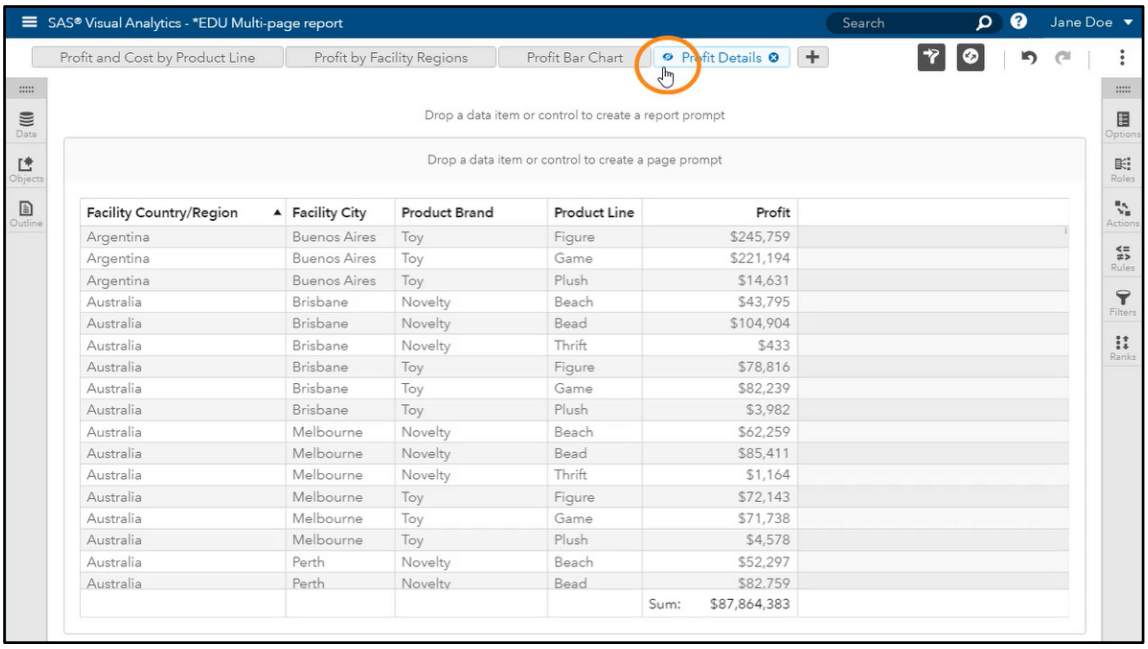


The third page, the Profit Bar Chart page shown in Figure 1.57, has the same measures as the previous page in a bar chart report object.

Figure 1.57: Profit Bar Chart

And the final page, the Profit Details page shown in Figure 1.58, displays profit data in a list table. You can see by the icon circled next to the page name that this is a hidden page.

Figure 1.58: Profit Details Page

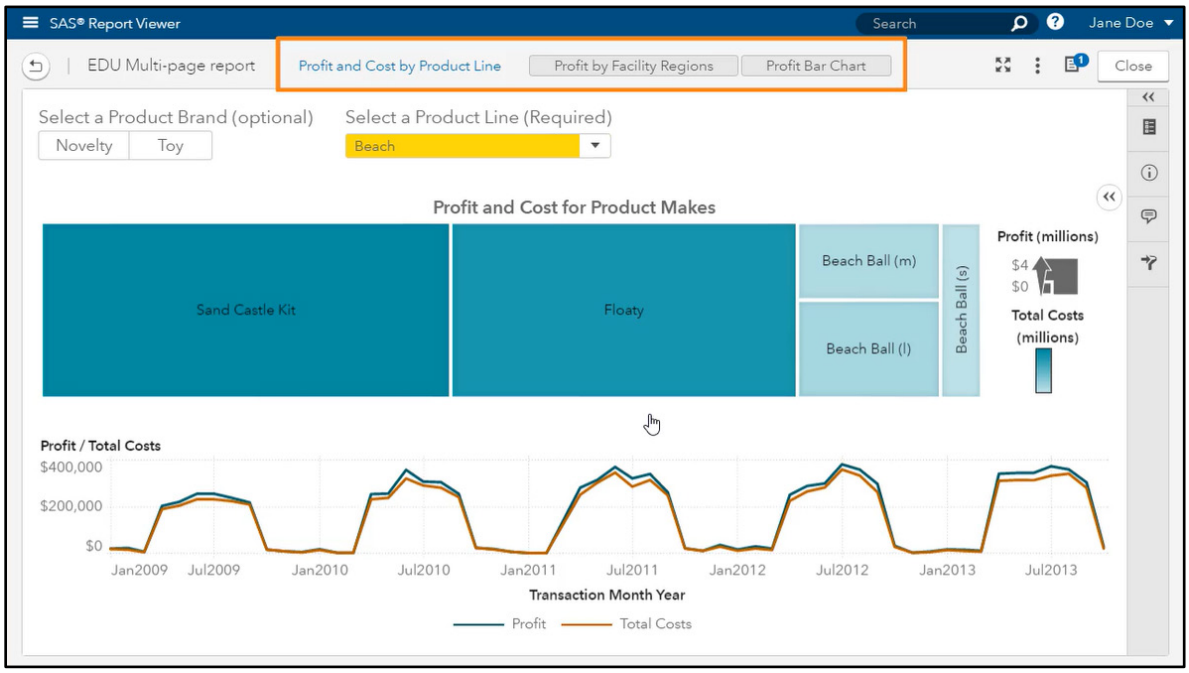


Using Hidden Pages

Now let's take a quick look at this report in the viewer. To go to the Report Viewer, select More Options (three vertical dots in the upper right-hand corner), and then select View Report from the drop-down menu.

In the Report Viewer, only three pages are displayed, as shown in Figure 1.59.

Figure 1.59: Report Viewer

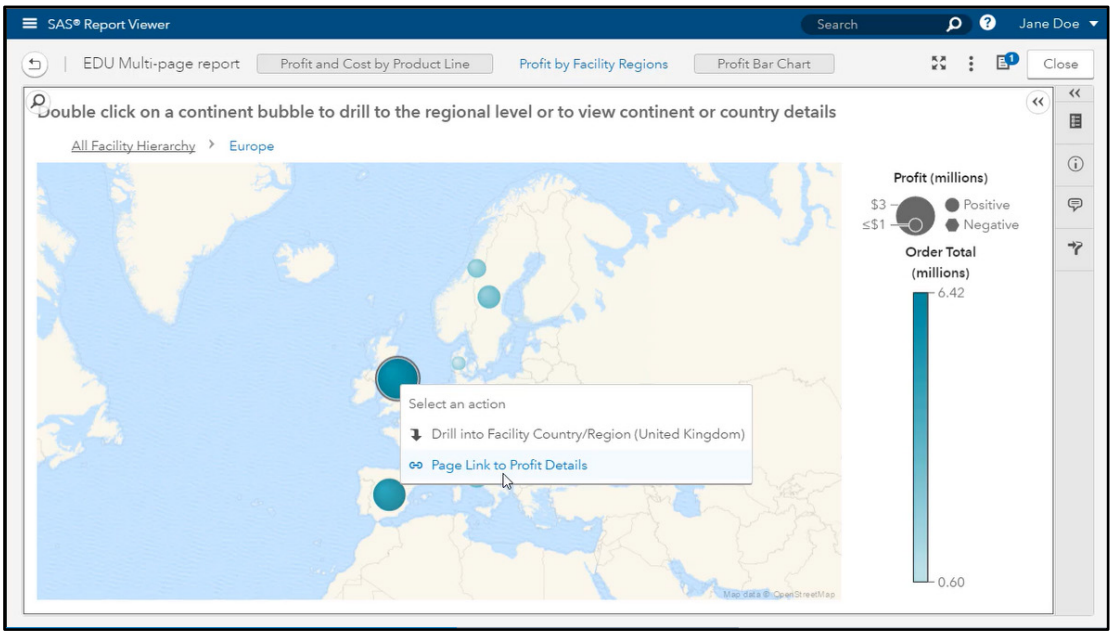


The Profit Details page is hidden, so it does not appear in the viewer. It is, however, available through a link from another section. Select the Profit by Facility Regions page to view the Geo map. As shown in Figure 1.60, the note indicates that you can double-click a bubble to show facilities at the next level of the hierarchy or view profit details. To start, double-click the bubble over Europe, and select the option to Drill into Facility Continents (Europe).

Figure 1.60: Profit by Facility Regions Drill Action

Now you can see additional bubbles that show Western European countries that have facilities. According to the size of the bubbles, the UK Facilities have the highest order totals and profit. Double-click the UK bubble as shown in Figure 1.61, and this time, select Page Link to Profit Details from the drop-down menu.

Figure 1.61: Profit by Facility Regions Link Action



The hidden page with the Profit Details table appears. This is a subset of the table that we saw on the hidden page in Visual Analytics. Click Maximize View to see the detail profit data in a larger version of the table as shown in Figure 1.62.

Figure 1.62: Hidden Page in Report Viewer

We can see that there are facilities in Manchester and London, and we see the profit for each product brand and line. Close this page, and then use the link to go back to the top of the hierarchy in the report.

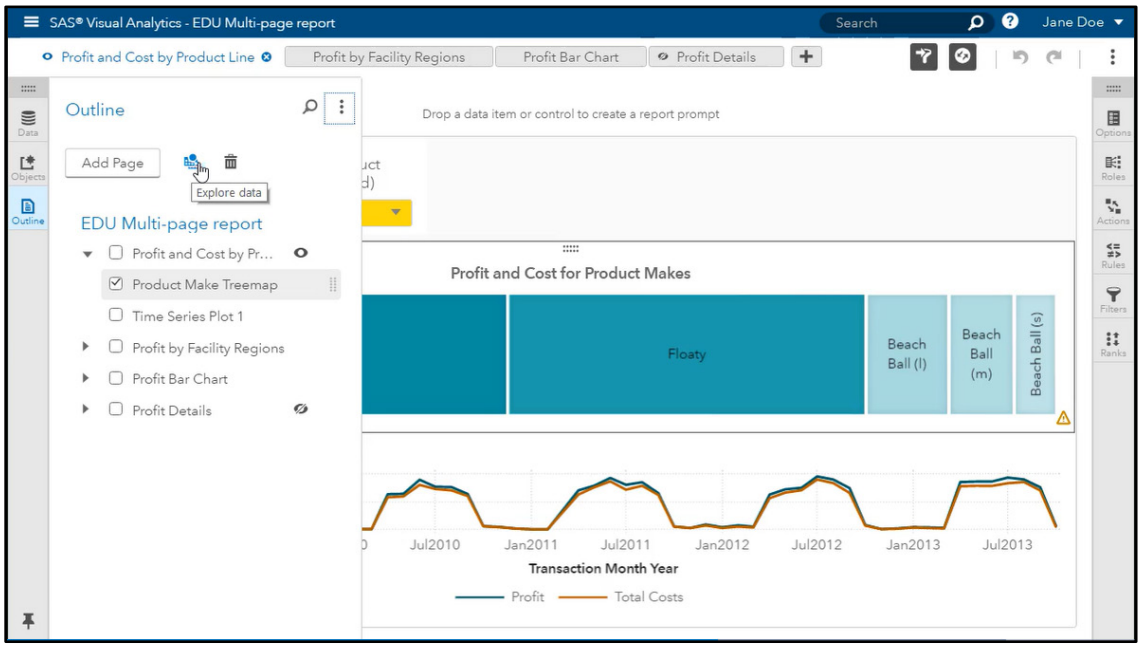
Rearranging Objects by Using the Outline Pane

In this section, we will continue looking at the example report from the previous section that has four pages with various analyses of profit for a toy company. Return to Visual Analytics to edit the report.

Recall that the second and third pages of the report show the same information displayed as two different chart types: a geo map and a bar chart. What if we want to fit both charts on the same page?

To move objects between pages, use the Outline pane on the left side of the screen. In the Outline pane there are several additional actions that you can take besides moving objects. You can add a page by using the Add Page button. Or you can select an object in the current report and use the buttons to either explore data or delete the object, as shown in Figure 1.63. You can also change the visibility of a page by clicking the eye icon on the right side of the page. Notice that you cannot change the visibility of the Profit by Facility Regions page. This page contains a link to the Profit Details page, so it cannot be hidden.

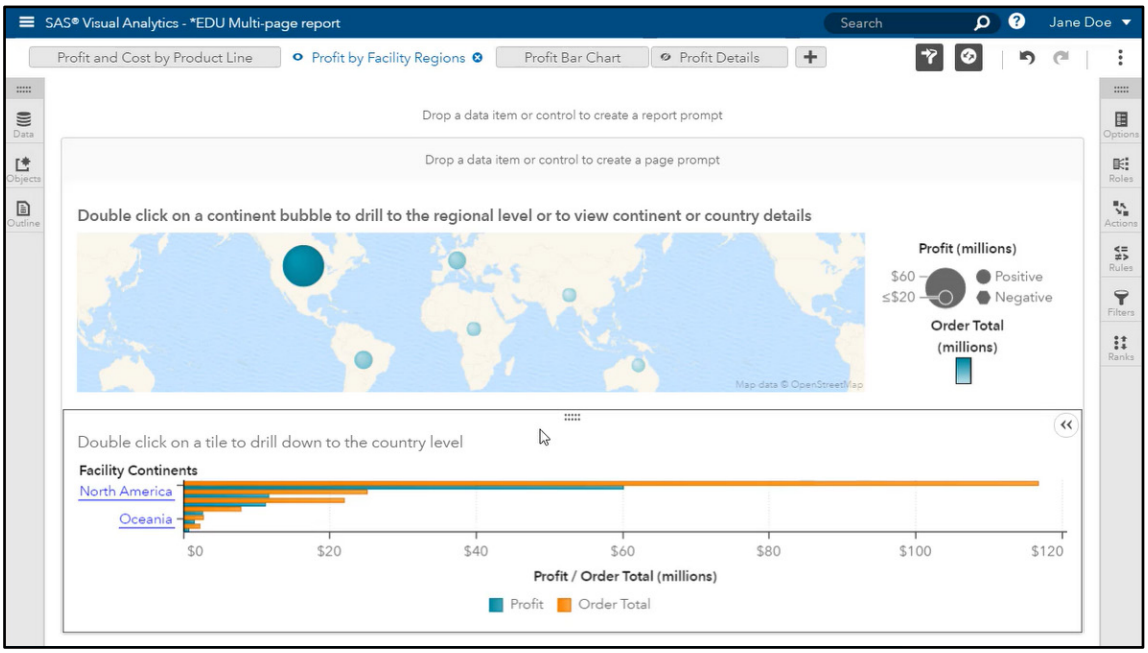
Figure 1.63: Outline Pane



Now let's see how to rearrange objects in a report. Expand the Profit by Facility Regions and Profit Bar Chart pages in the Outline pane. We want to move the bar chart on the Profit Bar Chart page to the Profit by Facility Regions page so that both the geo map and bar chart with the same information are together.

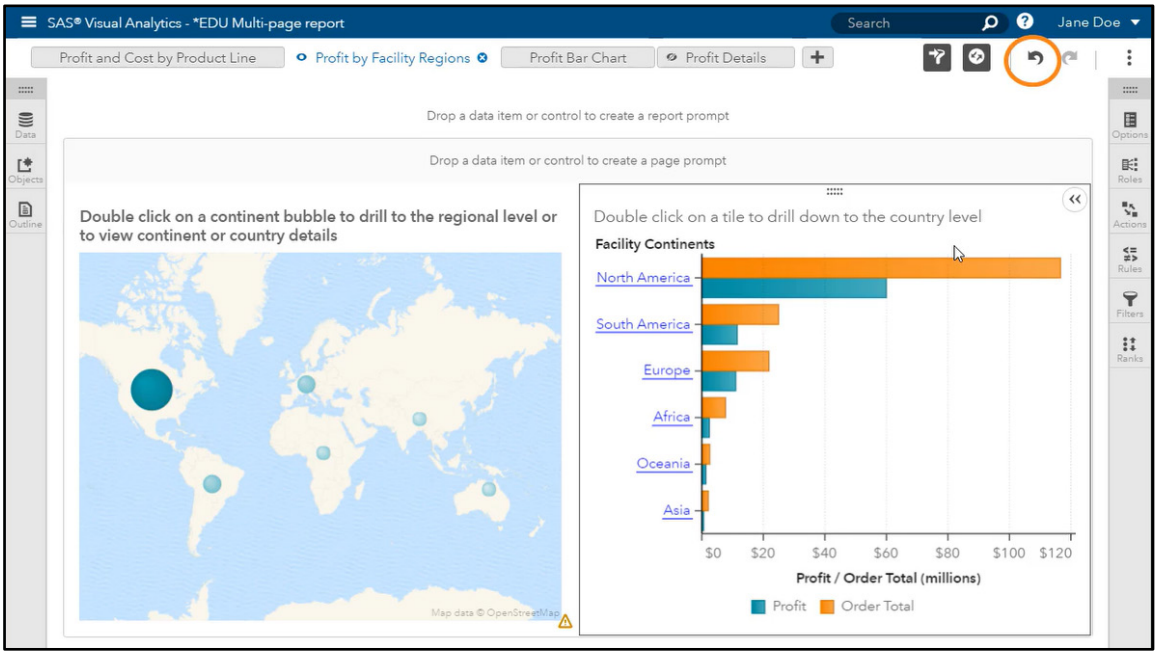
To move an object, click the Move button by moving your cursor over the double column of vertical dots to the right of the object that you want to move in the Outline pane. A four-sided arrow will appear. Drag the object to the new location. In this case, you will drag the bar chart to the Profit by Facility Regions page. The charts are now on the same page and stacked, as shown in Figure 1.64.

Figure 1.64: Stacked Charts



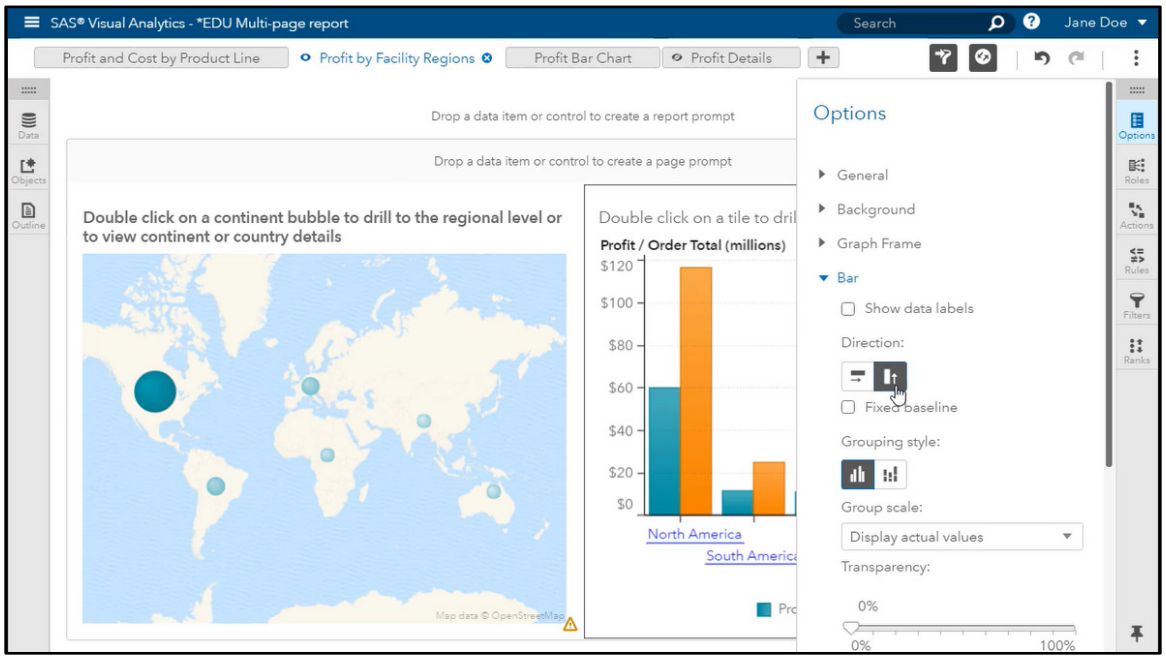
If you want to see the charts side-by-side instead of stacked, click the drag button at the top of the bar chart. It looks like a double row of horizontal dots. A four-sided arrow appears. Drag the object until the indicator line to the right of the geo map is reached. The charts are now side-by-side. Remember that, if you do not like a change, then you can use the Undo button in the upper right-hand corner of the window (circled in Figure 1.65).

Figure 1.65: Side-by-Side Charts



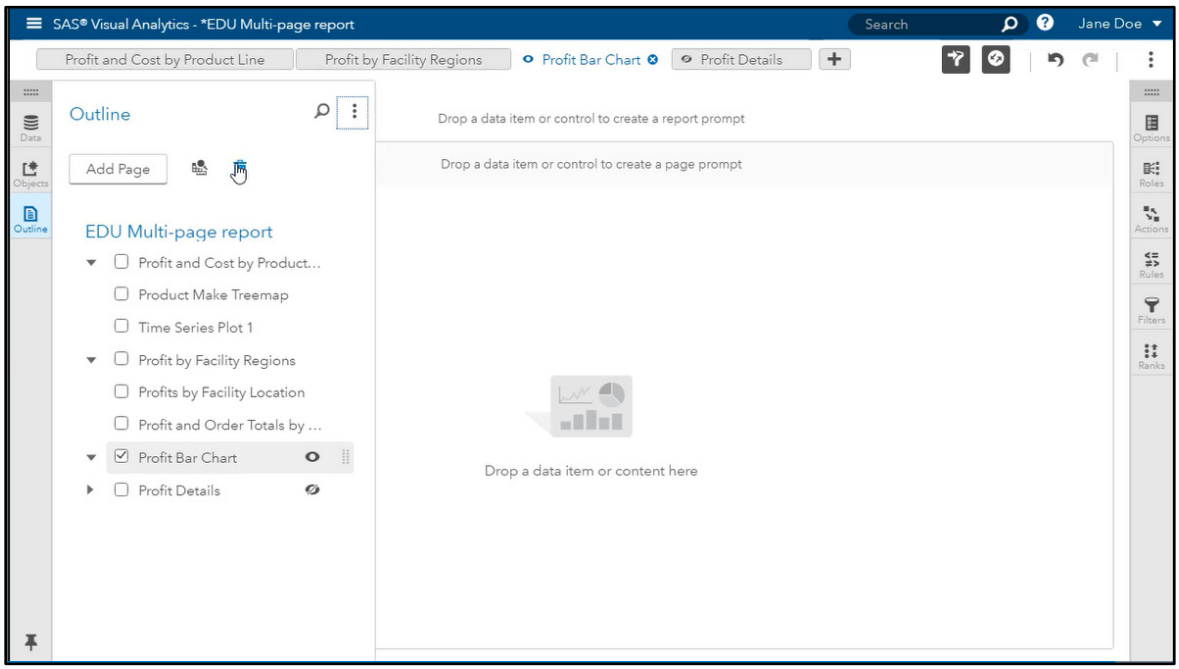
What if you prefer a vertical bar chart instead of a horizontal bar chart now? Make sure the bar chart object is selected, and then click Options from the rightmost pane. In the Bar section, you can change the direction to vertical, as shown in Figure 1.66. Then close the Options pane by clicking outside of it.

Figure 1.66: Options Pane



Because we moved the bar chart from the third page to the second page, the third page is now empty. We want to delete that page now. Go back to the Outline pane on the left side. Select the Profit Bar Chart page, and delete it by clicking on the trashcan icon, as shown in Figure 1.67.

Figure 1.67: Outline Pane Delete Page Option



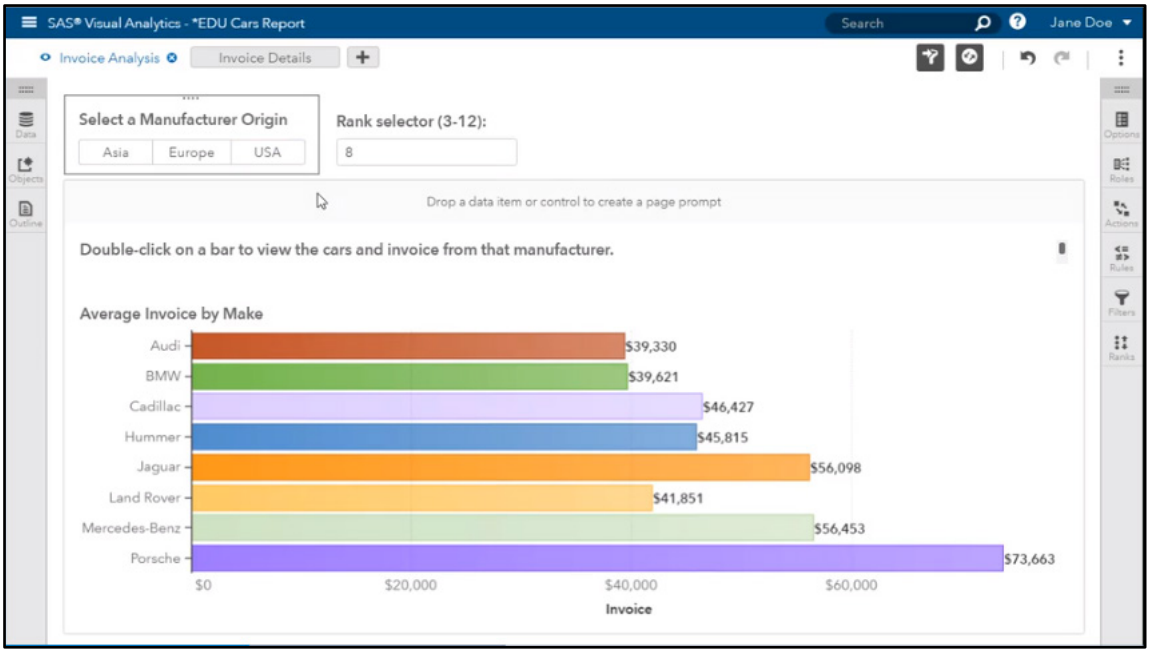
Your report now has three pages, one of which will be hidden in Report Viewer.

Creating New Charts Using Duplication and Changing Chart Types

Sometimes a convenient way to create or modify a SAS Visual Analytics report is to duplicate an existing chart and change the chart type. This enables you to quickly and easily view the same data in a different way.

To start, open an existing report in Visual Analytics. The report in this example has two pages. There is an optional report-level filter for manufacturer origin, and an input box for the rank selector. To see the top 8 worldwide manufacturers, enter 8 in the rank selector and do not choose a manufacturer origin, as shown in Figure 1.68.

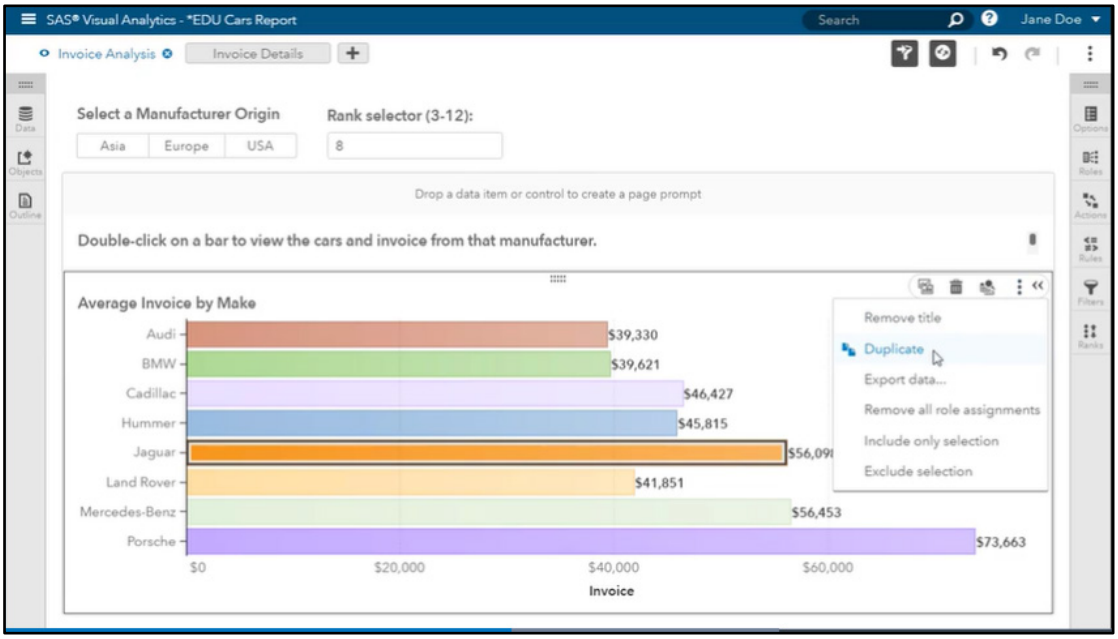
Figure 1.68: Invoice Analysis



What if some of your colleagues have asked to see additional summary statistics for manufacturer invoices on the Invoice Analysis page? A box plot seems like the best option to add the information that they want. We can add a new box plot to the report and then associate the roles, set properties, and add a rank to match the bar chart. But in this case, duplicating the bar chart will save many steps.

To begin the duplication process, open the Options menu for the bar chart, select More Options, and then Duplicate, as shown in Figure 1.69.

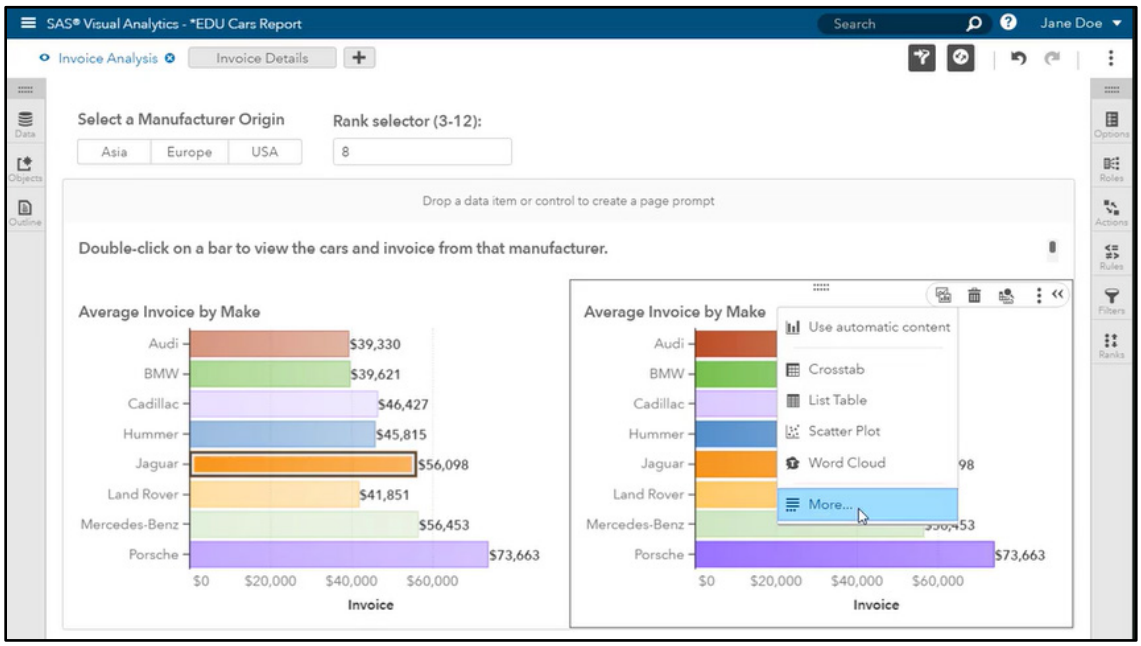
Figure 1.69: Duplicate Option



A copy of the bar chart is added below the original. All roles and other attributes, including the rank, are included in the new object. To make the two charts side-by-side, select the handle for the new chart and drag it to the right of the original bar chart.

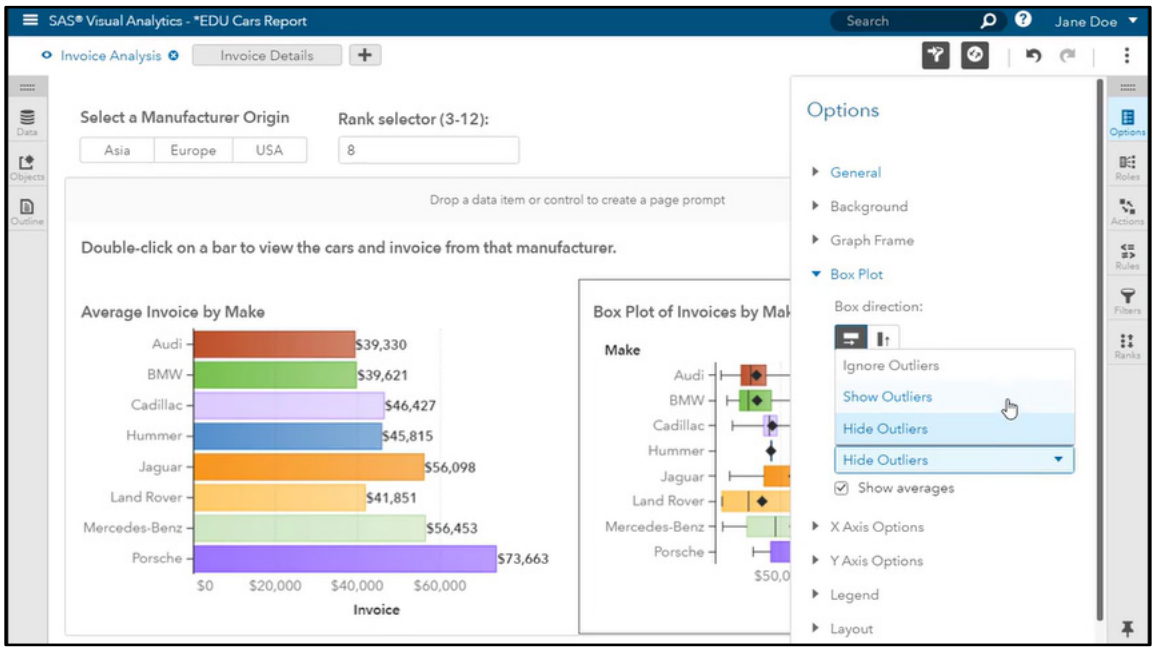
Now, we will change the chart type. In the Bar Chart Options menu, select Change Bar Chart To, and then More, as shown in Figure 1.70. In the Select a New Type window that pops up, select Box Plot.

Figure 1.70: Change Chart Type Options



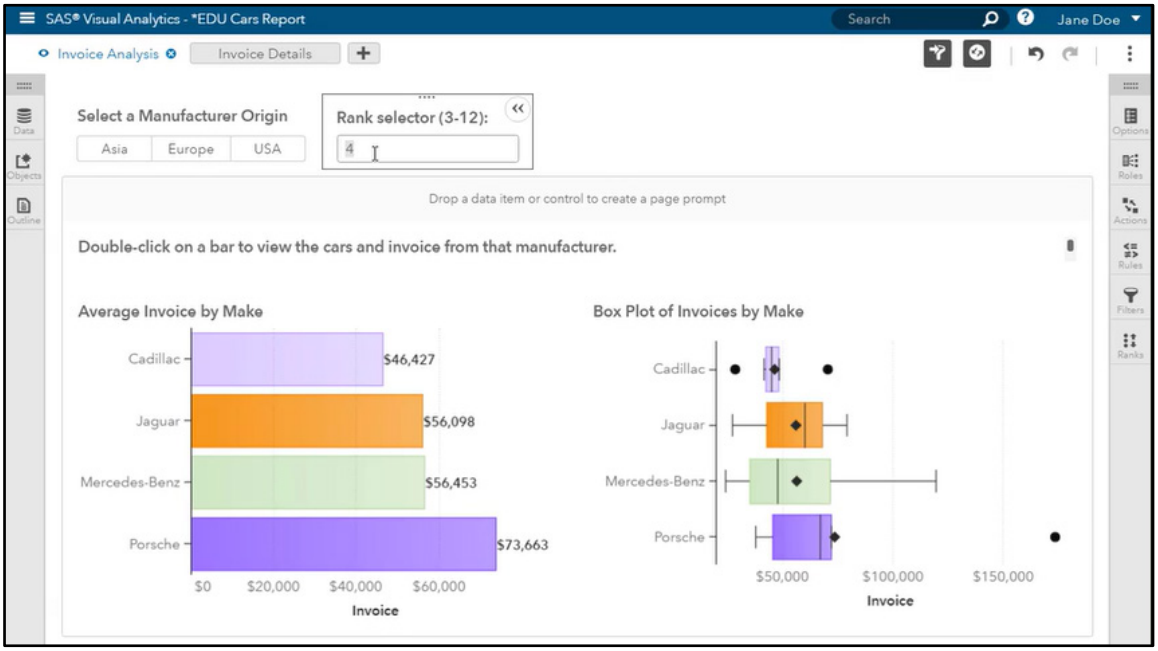
Now, let's modify some of the options for the Box Plot. Select Options in the rightmost pane, and under General, change the title to Box Plot of Invoices by Make. Under Box Plot, change the direction to horizontal to match the bar chart. Then select the Show Averages check box and, for Outliers, select Show Outliers, as shown in Figure 1.71. Under Y Axis options, clear the Axis Label check box.

Figure 1.71: Options Pane



Close the Options pane by clicking outside of it. Let's look at the top 4 worldwide manufacturers. Enter the number 4 in the Rank Selector box, and both charts are updated, as shown in Figure 1.72.

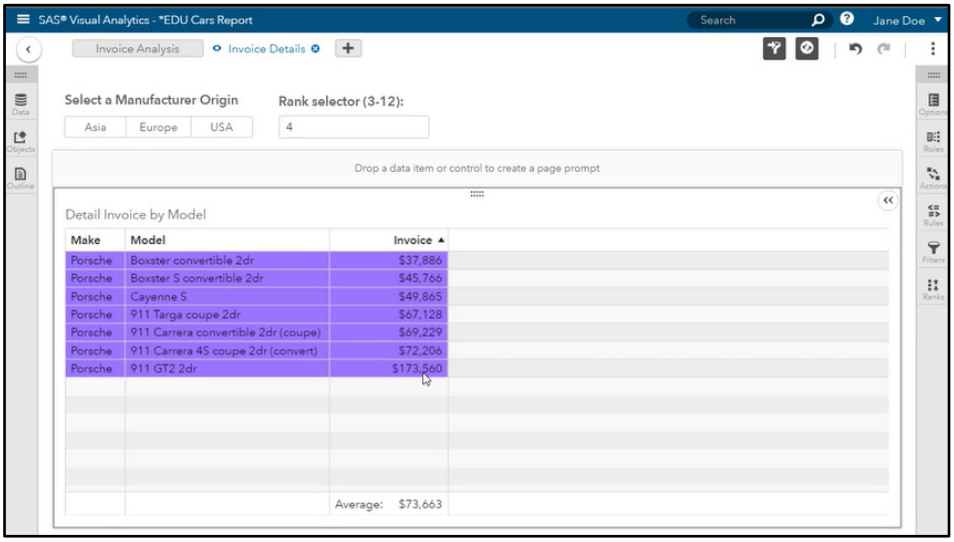
Figure 1.72: Modified Invoice Analysis Page



As you can see, the box plot gives much more information about the invoices for each manufacturer. The vertical line in each box represents the median, and the diamond represents the average. For most makes, the average value is greater than the median. If you place the mouse on the diamond in the Mercedes-Benz box, you will see a pop-up data tip with a lot of information about invoices for this manufacturer. For Cadillac, it is easy to see that the average is close to the median, but there are 2 outliers: one greater than and one less than the expected values. Another interesting make is Porsche.

Porsche is the manufacturer with the highest overall average, but that is due to one outlier that pulls the average higher than expected. Let's see what model is the outlier for Porsche. Double-click the Porsche bar in the bar chart on the left. This opens the Invoice Details page, as shown in Figure 1.73. In the list table, we see that the model is the 911 GT2 2dr.

Figure 1.73: Invoice Details Page



Resources

This chapter is based on the "SAS Visual Analytics on SAS Viya" videos in SAS® Viya® Enablement, a free course available from SAS Education.

You might find the following documentation and resources helpful as you learn more about programming in SAS Visual Analytics:

- SAS Visual Analytics documentation
- Insight Toy sample data available to download in SAS Visual Analytics 8.1 documentation
- SAS Visual Analytics 8.3: Programming Guide
- SAS Technical Papers
- SAS Viya training courses from SAS Education
- SAS Visual Analytics Communities
- *An Introduction to SAS® Visual Analytics: How to Explore Numbers, Design Reports, and Gain Insight into Your Data* by Tricia Aanderud, Rob Collum, and Ryan Kumpfmiller

Chapter 2: SAS Visual Statistics

Introduction **43**
Getting Started with SAS Visual Statistics **43**
Clustering Model **45**
Parallel Coordinates Plot 46
Cluster Matrix 48
Actions 48
Decision Tree Model **49**
Model Assessment Visualizations 52
Actions 54
Logistic Regression Model **54**
Missing Values 56
Model Assessment Visualizations 56
Group-by Role 57
Actions 58
Resources **59**

Introduction

SAS Visual Statistics 8.2 on SAS Viya is a web-based solution designed for statisticians, data scientists, and business analysts who want to interact with and analyze complex data visually and instantly. The easy-to-use drag-and-drop interface enables you to create descriptive and predictive models on data of almost any size, without the need to write code. The new HTML5-based interface makes it easier than ever to combine data discovery, predictive modeling, and interactive reporting.

SAS Visual Statistics runs on SAS Viya, the innovative, open platform that is designed for any analytical challenge. SAS Viya maximizes parallelization and distribution of analytical workloads to give you high-impact insights faster. This new engine provides nearly instantaneous results and enables multiple users to have concurrent access to the same in-memory data, with no conflict.

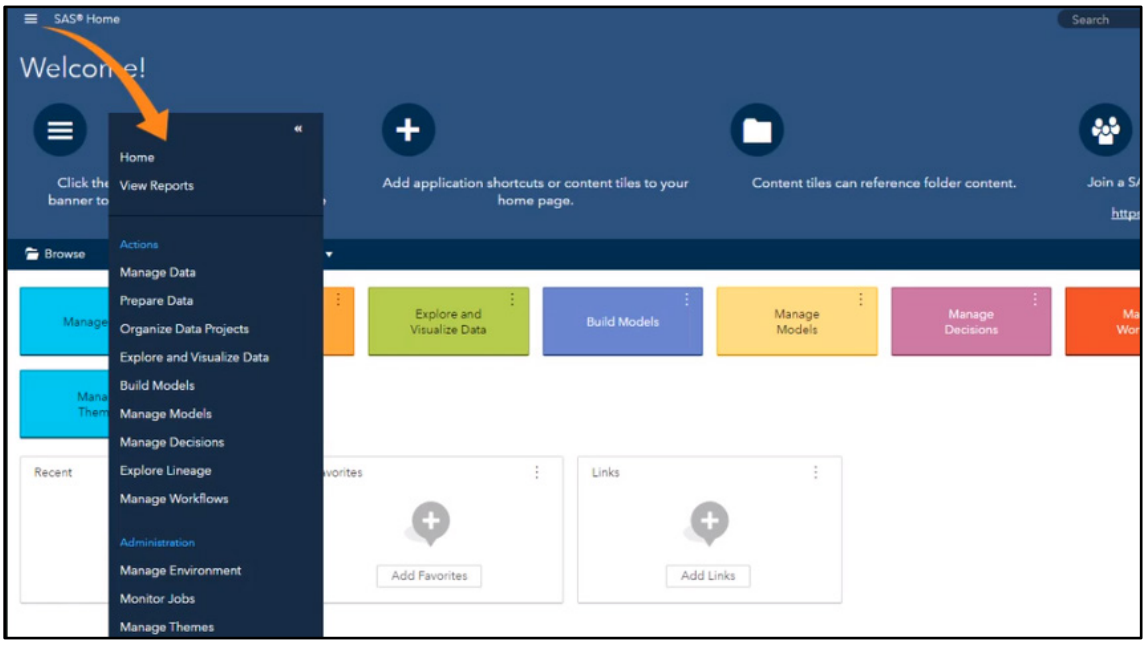
SAS Visual Statistics is an integrated add-on to SAS Visual Analytics. Combining the two products gives you a single, fast environment for interactive data exploration and model development. With Visual Statistics, you use statistical or machine learning to do the following:

- Recognize patterns and assess likely outcomes
- Shorten model development time by taking advantage of group-by refinement
- Identify meaningful segments of your data
- Assess model performance by establishing training and validation partitions

Getting Started with SAS Visual Statistics

There are two ways to get started working in SAS Visual Statistics. From the SAS Home page, you can select the Explore and Visualize Data tile in the main window. Or, if you are using another SAS Viya application, you can click the application menu bar in the upper left-hand corner. Then, using the side menu, click Explore and Visualize Data, as shown in Figure 2.1.

Figure 2.1: SAS Home Page and Application Menu Bar



Some of the features within the SAS Visual Statistics interface that make it easy for you to build your models are shown in Figure 2.2.

Figure 2.2: Visual Statistics Features

❶ The **Application Bar** provides access to the Home Page and other SAS Applications.
❷ The **Menu Bar** offers common tasks, such as creating pages, undo, and redo.
❸ The **Left Pane** enables you to work with your data and select your model object type.
❹ The **Canvas** is the workspace for building one or more models.
❺ The **Right Pane** enables you to set model roles, options, and filters.

After you add a data source, you can access the SAS Visual Statistics models from the Objects pane accessed from the left pane. Simply double-click or drag and drop the model of your choice onto the canvas to get started.

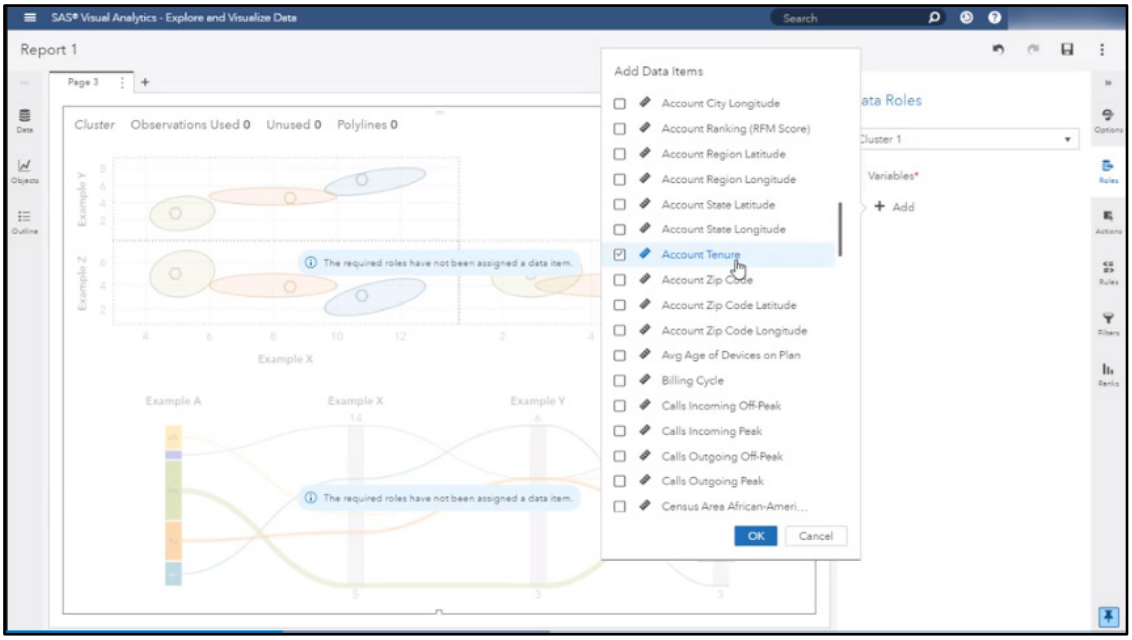
Clustering Model

In this section, you will learn how to use SAS Visual Statistics to build a customer segmentation model. We will use the clustering model within SAS Visual Statistics to perform data-driven segmentation. In this example, we will use telecommunications customer attrition data.

Suppose we want to understand a little more about the telecommunication customer accounts, based on tenure and the lifetime value of the accounts, as well as the estimated age and house value of the area where an account holder lives. We then want to segment customers on the basis of this information.

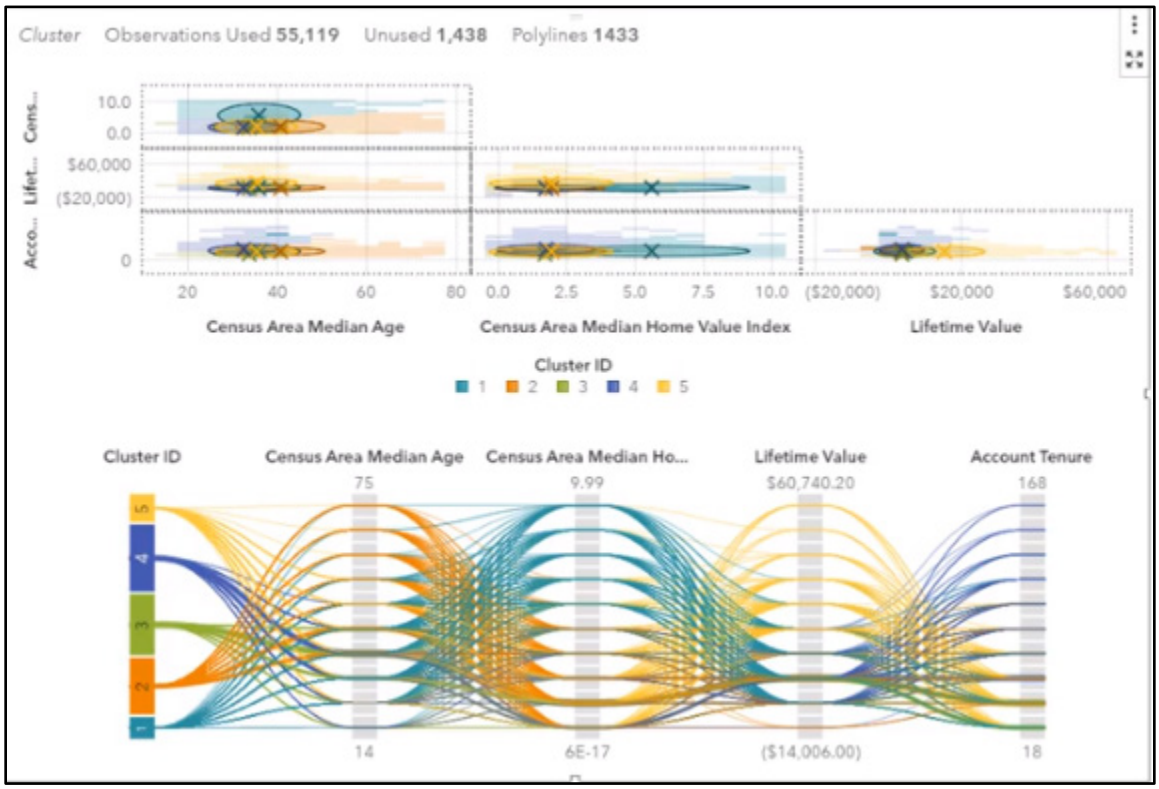
Once you have loaded your data, you will create a new clustering visualization. Click the Objects option in the far-left pane. From the SAS Visual Statistics category, click Cluster, and drag it into the canvas area. Then, in the far-right pane, click Roles. Assign the following four attributes to the analysis, as shown in Figure 2.3: Account Tenure, Census Area Median Age, Census Area Median Home Value, and Lifetime Value. Click OK.

Figure 2.3: Add Variables



SAS Visual Statistics uses a *K*-means clustering approach. By default, it generates five segments. The output in the canvas area is shown in Figure 2.4.

Figure 2.4: Cluster Results

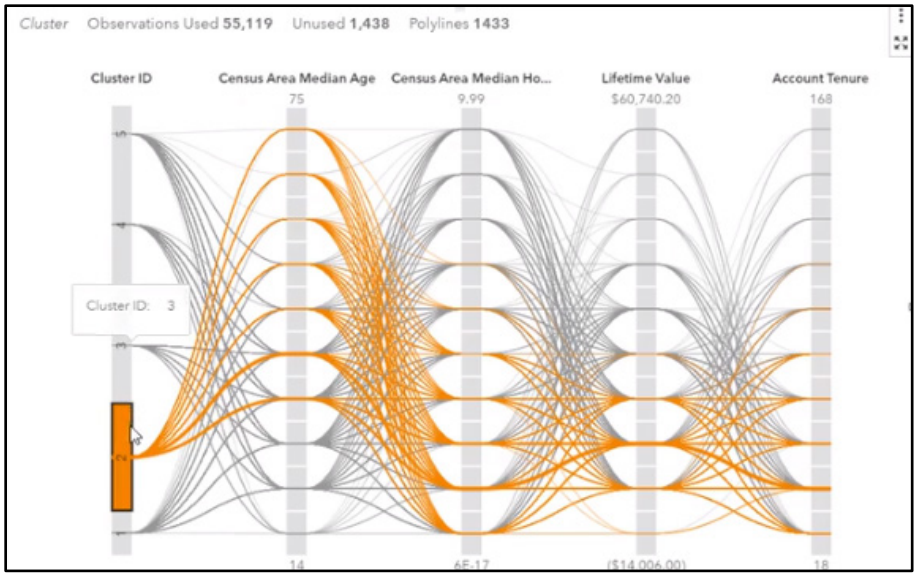


Parallel Coordinates Plot

The visualization at the bottom of Figure 2.4 is a parallel coordinates plot. It shows the five cluster IDs that were generated as the output of the clustering analysis. On the left vertical line and across the top horizontal line, we are looking at the attributes for Census Area Median Age, Census Area Median Home Value, Lifetime Value of the Account, and Account Tenure.

If you want to explore each of the segments, simply click one of the cluster IDs on the left, and it will isolate that segment, as shown in Figure 2.5.

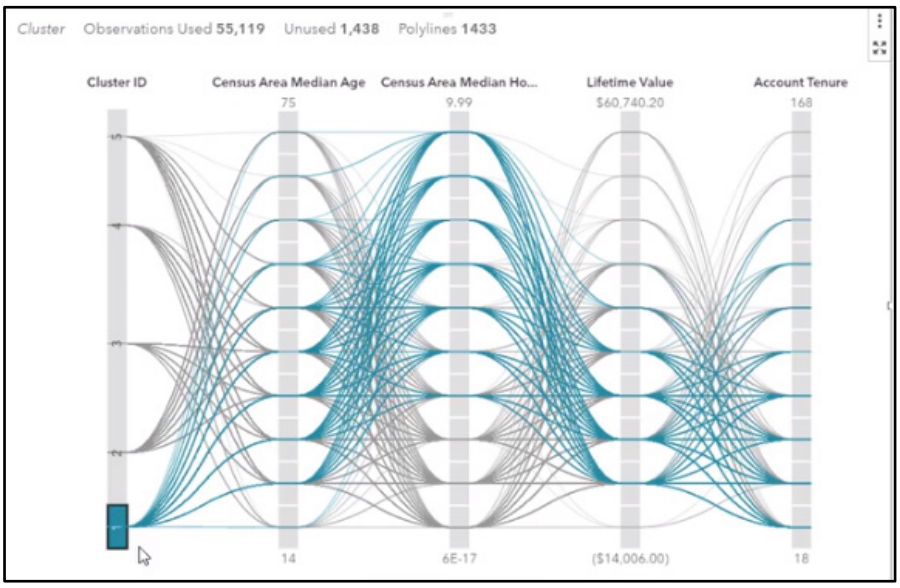
Figure 2.5: Cluster ID 2 Segment



In Figure 2.5, we see cluster 2, or segment 2. We can see that the estimated age of the account holders was high, and the estimated home value of the account holders was low. The accounts were relatively new, according to Account Tenure, and had low lifetime value.

Now, let's look at a different segment or cluster. Click Cluster ID 1 in the left vertical bar to see the visualization shown in Figure 2.6.

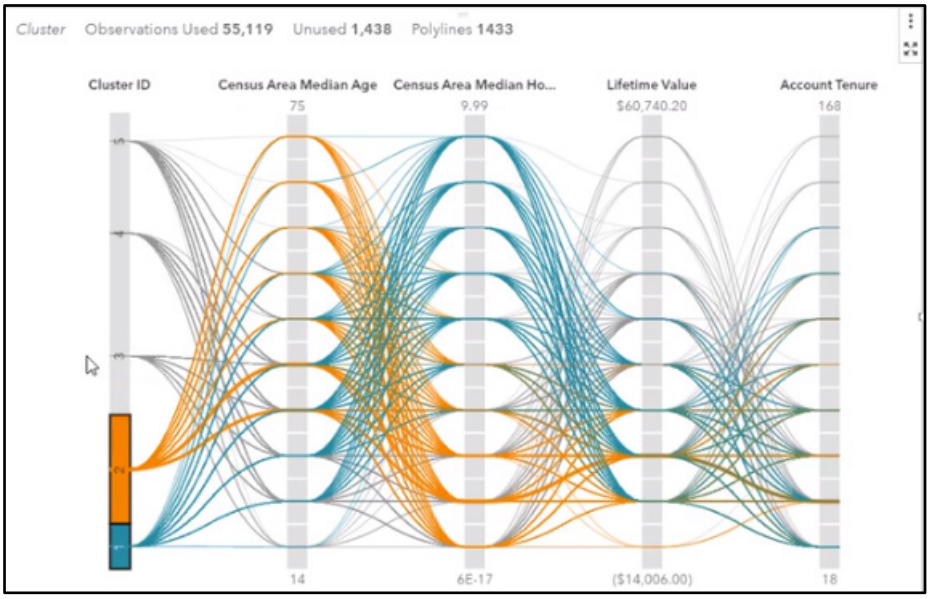
Figure 2.6: Cluster ID 1 Segment



Here we see a different story. Compared with segment 2, these account holders were estimated to be younger and with higher house values. Their account tenure was longer. Therefore, the lifetime value of these accounts was higher. It looks like this is a good subset of loyal customers of high account value. The next step could be a study of this subset to avoid "churning."

If you want to compare patterns across segments, simply hold down the Ctrl key and click to select the previous cluster, as shown in Figure 2.7.

Figure 2.7: Cluster ID 1 and Cluster ID 2 Segments

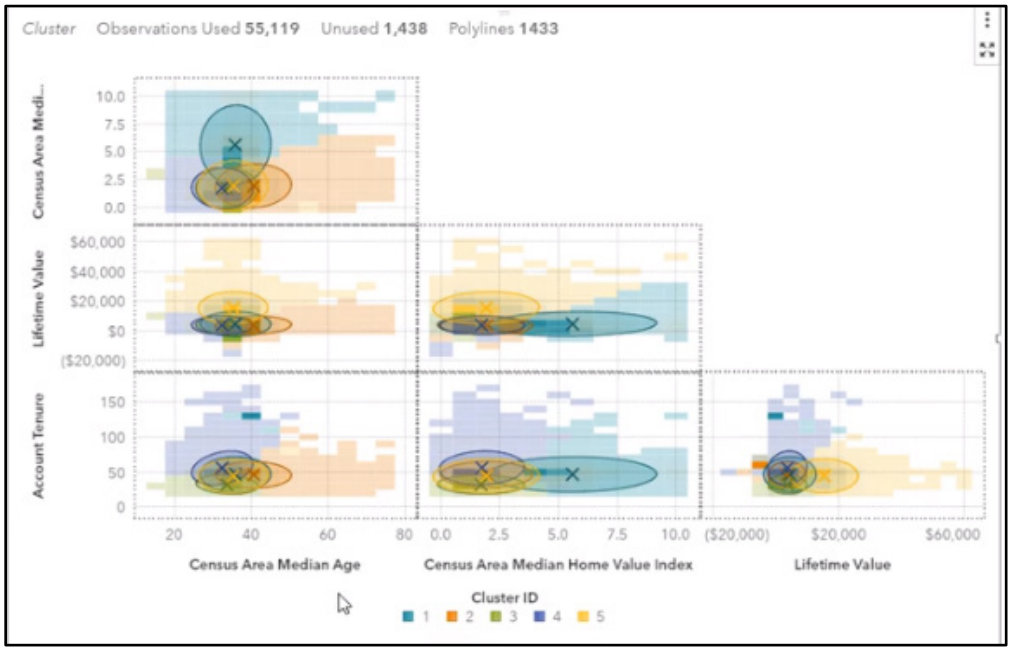


You can see that, by using the clustering analysis, we can quickly identify meaningful customer segmentations.

Cluster Matrix

To go further in our analysis, we can move over to the cluster matrix, which was the top visualization in the original report. The cluster matrix gives us an expanded view through visualization, as shown in Figure 2.8. We can see how much overlap we have between the clusters for each pair of the four attributes that we folded into the analysis.

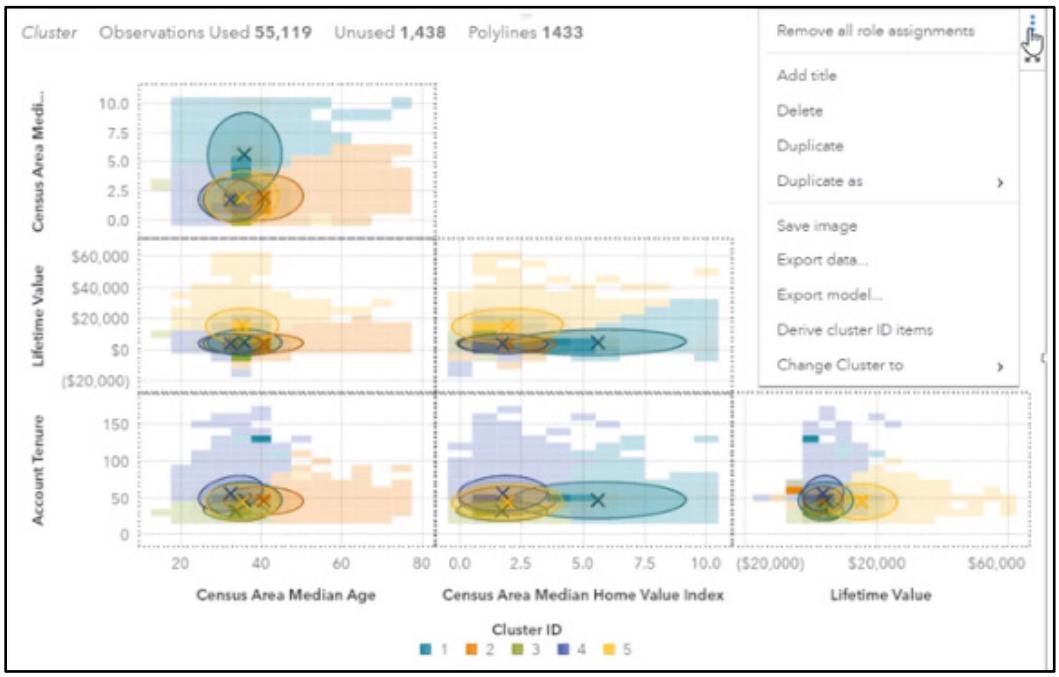
Figure 2.8: Cluster Matrix



Actions

Next, let's explore how we can take action on the results of this analysis. By clicking the drop-down button in the upper right corner of the canvas, you are given several options, as shown in Figure 2.9.

Figure 2.9: Cluster Actions



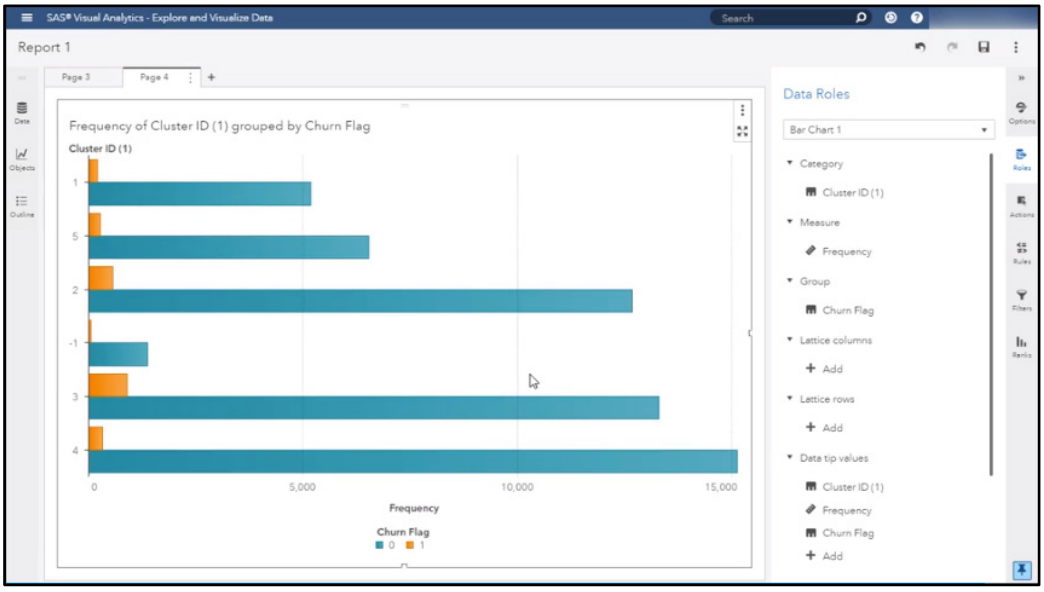
Export Model

Clicking Export Model will export the auto-generated score code of this analysis as a SAS DATA step. You could run the score code in real time to allocate new customers to one of the five segments in accordance with their census and account values.

Derive Cluster ID Items

In addition to exporting score code, you can derive a cluster ID variable by clicking on Derive Cluster ID Items. SAS Visual Statistics generates a new Cluster ID column that assigns each customer to one of the five segments. This will add a new column to your data, which you can see if you look at the Data Option in the far-left panel. You can then use this column in a subsequent analysis to compare attrition rates across the five customer segments, as shown in Figure 2.10. This capability is very valuable because it enables you to enrich your existing data set with new information.

Figure 2.10: Cluster ID Frequency Bar Chart



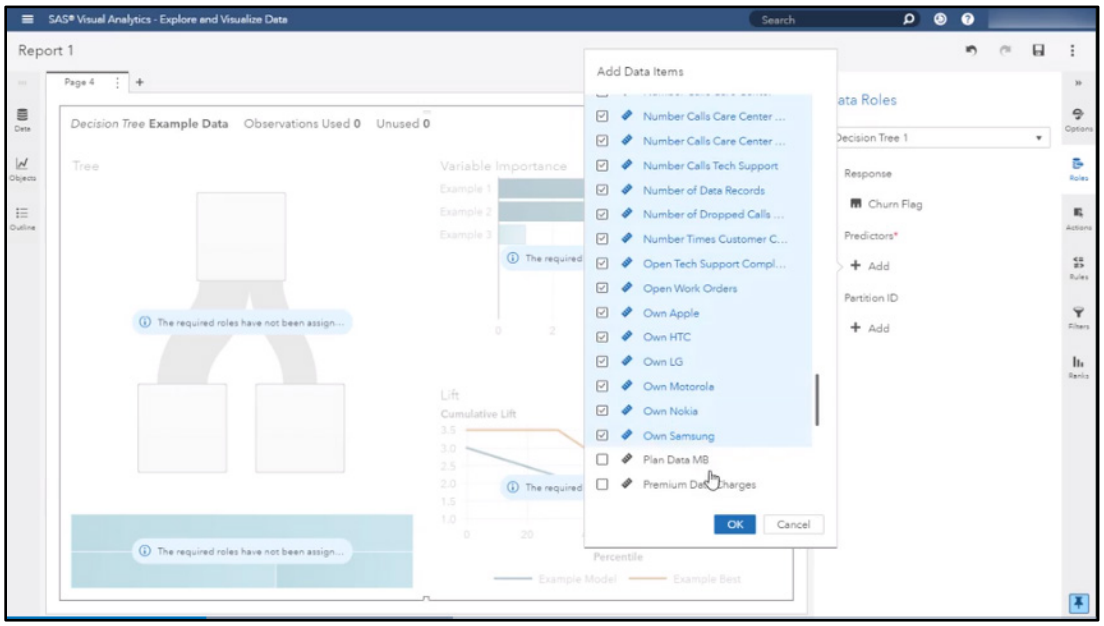
Decision Tree Model

In this section, we will learn how to use SAS Visual Statistics to build a decision tree model to study telecommunication customer data. In this example, we will identify important attributes that are related to whether customers of a telecom company cancel their service or close their account. This kind of behavior is often called *churn* or *attrition*.

First, we will create a new decision tree visualization once your data is loaded. Click Objects in the far-left pane. Under SAS Visual Statistics, choose Decision Tree and drag it into the center canvas. In the far-right pane, open the Roles panel. The target variable of the decision tree analysis is the Churn Flag. It is a binary variable that indicates whether the account is active or closed. Under Response, click Add and choose Churn Flag.

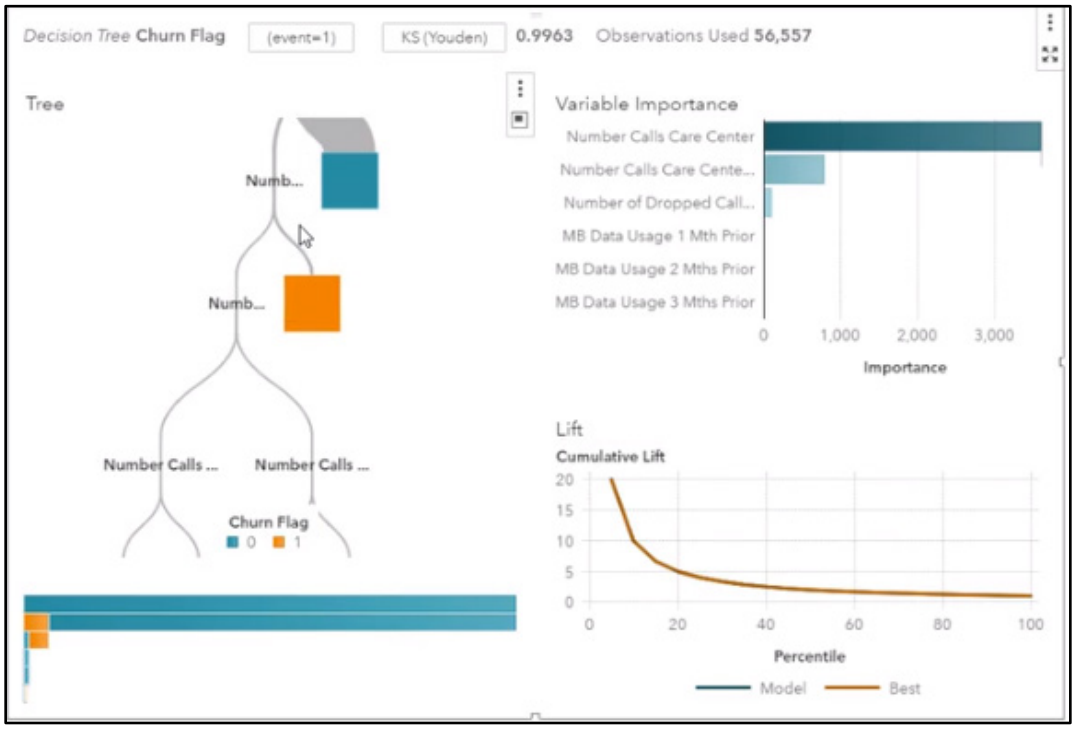
Next, we will select the attributes that we want to add to our analysis. Under Predictors, click Add, and choose all variables between MB Data Usage 1 Mth Prior and Own Samsung, as shown in Figure 2.11. These attributes are mainly related to recent account activity, such as data usage and minutes, recent calls to the care center and tech support, complaints, open work orders, and so on. Click OK.

Figure 2.11: Add Predictors



Notice that SAS Visual Statistics quickly grows and prunes a decision tree. It also identifies which attributes are important in general and which attributes are not important to customer churn. The results are shown in the canvas in Figure 2.12.

Figure 2.12: Decision Tree Results



You can click and press on the decision tree to move it around. Let's zoom in and look at the first node in the decision tree in Figure 2.13. This node represents all the data.

Figure 2.13: Decision Tree Node 0 Detail

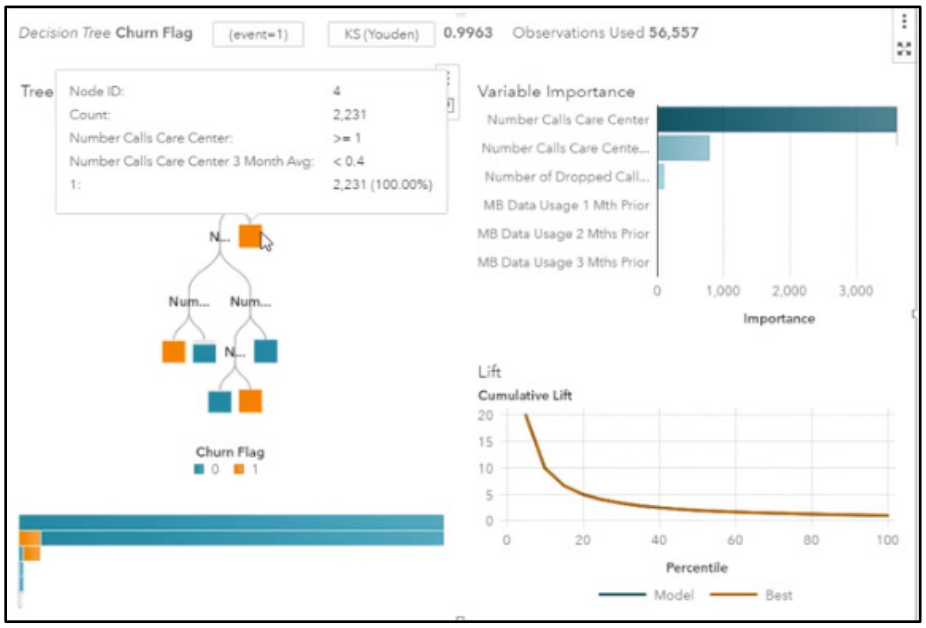
Approximately 4% of the customers canceled their service. The attribute that was most important in partitioning the data was Number Calls Cares Center. We can see that, if a customer has never called the Care Center, then he or she is unlikely to cancel service. The churn rate is actually 0%.

In contrast, 2,960 customers had at least one call to the Care Center, and the average churn rate is 79.53%. This is why the predictor Number Calls Care Center was selected as the most important attribute of customer churn. It separates a large portion of customers who were unlikely to cancel service from a relatively small subset of customers who had a 79.53% likelihood of canceling service.

SAS Visual Statistics uses the C4.5 algorithm and the information gain criterion to select the attribute that is used to split the data recursively and then build a decision tree.

Now let's zoom out and look at the entire decision tree and hover over Node 4, as shown in Figure 2.14.

Figure 2.14: Decision Tree Node 4 Detail



Node 4 contains 2,231 canceled accounts and no active accounts. These observations have Number Calls Care Center equal to or greater than 1, but recent 3-month average number of calls less than 0.4. This might indicate a subgroup of unhappy customers with relatively long account age. The decision tree tells us that these types of customers have all canceled their service.

In general, we can see that the decision tree model helps us segment the customers into subgroups, each with a different likelihood of churn. We can also see that for each customer subgroup, the decision tree provides a business rule to help us characterize the group.

Model Assessment Visualizations

If you want to go further in your analysis, SAS Visual Statistics provides several model assessment visualizations, such as the Lift chart, the ROC curve, and the Misclassification chart. To view any of these visualizations that are not already included in the report, right-click a chart, and then choose other visualizations from the drop-down list, as shown in Figure 2.15.

Figure 2.15: Options Menu

Misclassification Chart

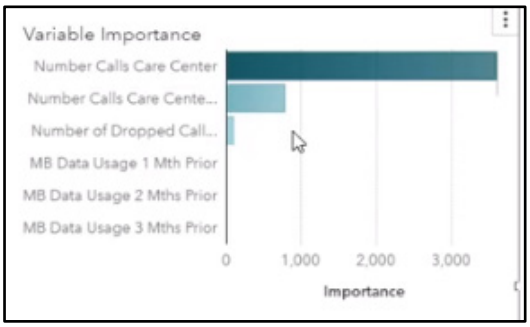
Click Misclassification in the drop-down menu to replace the Lift chart with the Misclassification chart. The Misclassification chart enables you to evaluate the quality of your model, because you can see how many observations the model is correctly and incorrectly predicting, as shown in Figure 2.16.

Figure 2.16: Misclassification chart

Variable Importance Plot

The Variable Importance plot ranks the attributes on the basis of their contribution to the splits in the entire tree. If an attribute is used in an earlier split, it contributes more to the target variable. In this example, Number Calls Care Center and Number Calls Care Center 3 Month Avg are the two most important variables, as shown in Figure 2.17.

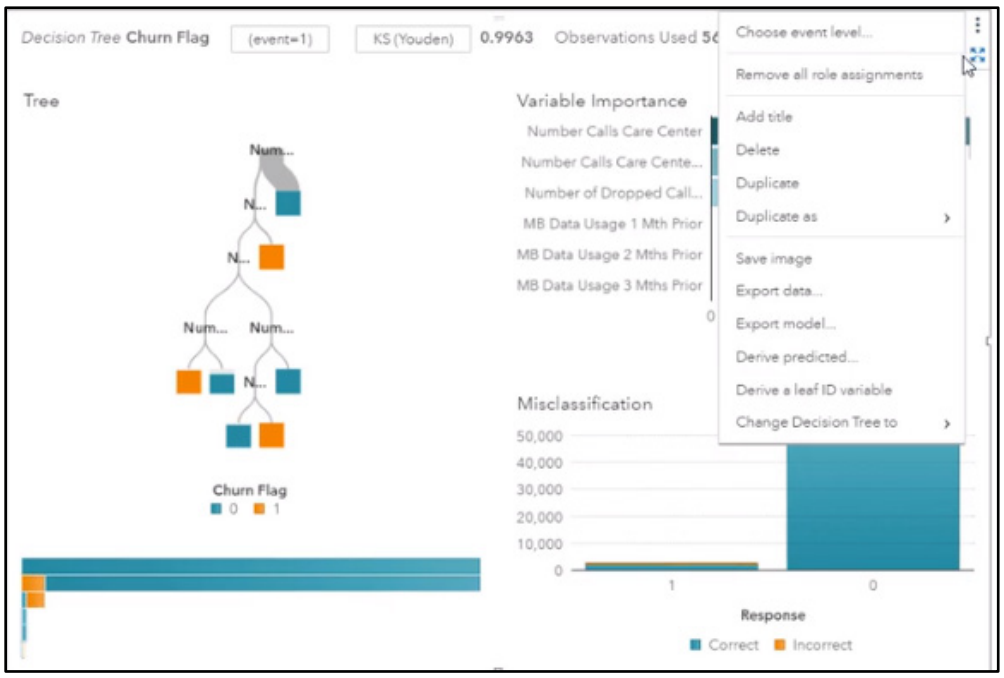
Figure 2.17: Variable Importance Plot



Actions

How do we take action on this decision tree? There are a couple of ways. Click the drop-down button in the upper right corner of the canvas to see all actions, as shown in Figure 2.18.

Figure 2.18: Decision Tree Actions



Export Model

Clicking Export Model will export the score code. The *score code* is a SAS DATA step that represents the decision tree model. You can take the intelligence of this tree and use it to score a new customer, identify the possibility of churn for that customer, and deliver that information to your peers in marketing.

Export Data

Clicking Export Data will export the model as a Microsoft Excel spreadsheet.

Derive a Leaf ID Variable

The Leaf ID variable assigns each customer to one of the terminal nodes in the decision tree. Clicking Derive a Leaf ID Variable automatically generates this variable without copying the data, and you can use it in further data analysis or in a business report.

Logistic Regression Model

In this section, you will learn how to use SAS Visual Statistics to build a logistic regression model to analyze customer churn data. We will use the model to explore what drives customers of a telecommunications company to cancel their services and leave for competitors.

After you have loaded the data, we will create a new logistic regression. Click Objects in the far-left pane. Under SAS Visual Statistics, choose Logistic Regression, and drag it into the center canvas. In the far-right pane, open the Roles panel. The target variable of the logistic regression analysis is Churn Flag. It is a binary variable that indicates whether a customer has canceled service. Under Response, click Add and choose Churn Flag.

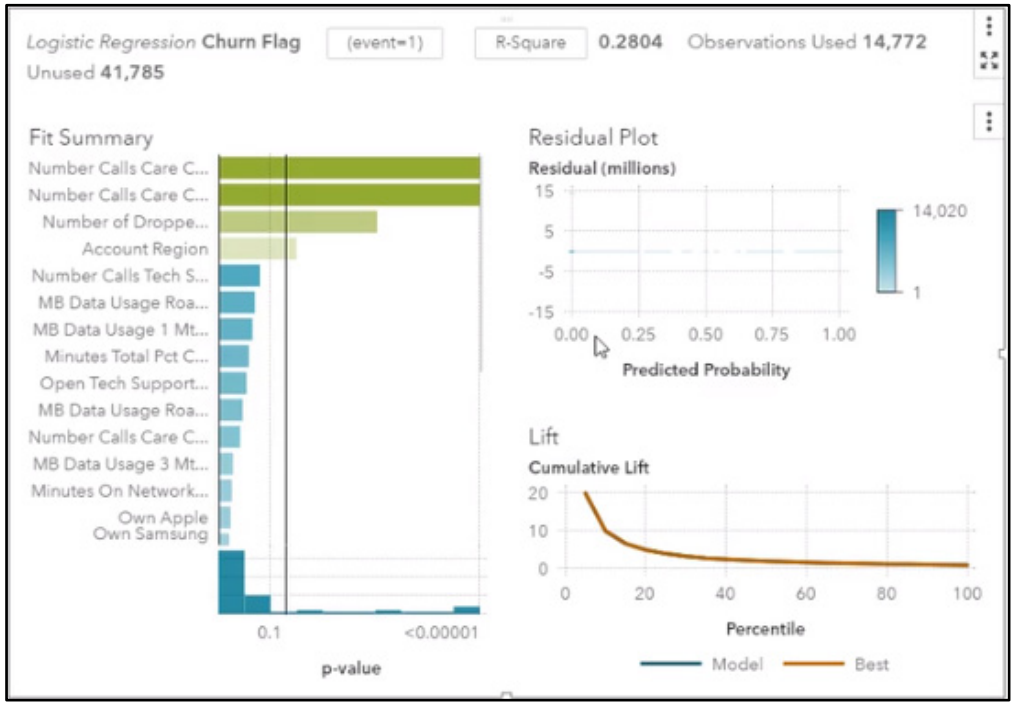
Next, we will select certain attributes that we want to add into the analysis. Under Continuous Effects, click Add. Choose all variables between MB Data Usage 1 Mth Prior and Own Samsung. These selected attributes listed in the Add Data Items menu are mainly related to the recent status of the account. For example, does the

customer have a data plan, and what is the recent data usage? Are there any calls to the customer care center and tech support center? Are there any open complaints or work orders?

We also want to find out whether certain regions have a higher percentage of customer churn than others. Under Classification Effects, click Add and choose Account Region.

The completed model is shown in Figure 2.19.

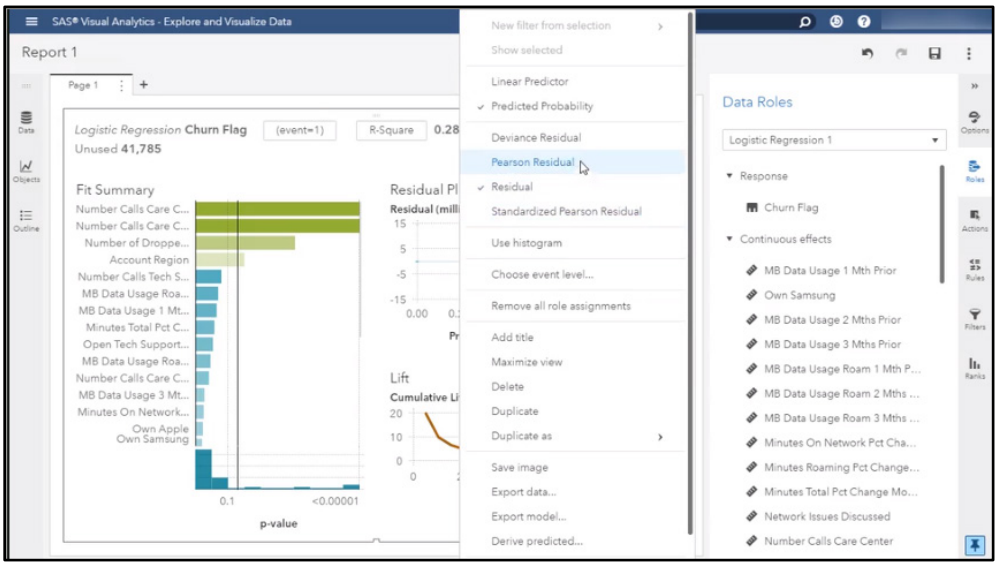
Figure 2.19: Logistic Regression Results



The first visualization on the left is the Variable Importance plot, or Fit Summary. This plot displays the negative log value of each variable's *p*-value. A Green bar indicates that the predictor is significantly related to customer attrition. Blue bars indicate insignificant predictors with a *p*-value greater than 0.05.

The output in Figure 2.19 also includes a Residual Plot for model diagnosis. This plot enables you to quickly identify any possible outliers in the analysis. You can right-click the statistic name as shown in Figure 2.20 to switch to other types of residuals, such as Deviance Residual, Pearson Residual, and Standardized Pearson Residual.

Figure 2.20: Options Menu

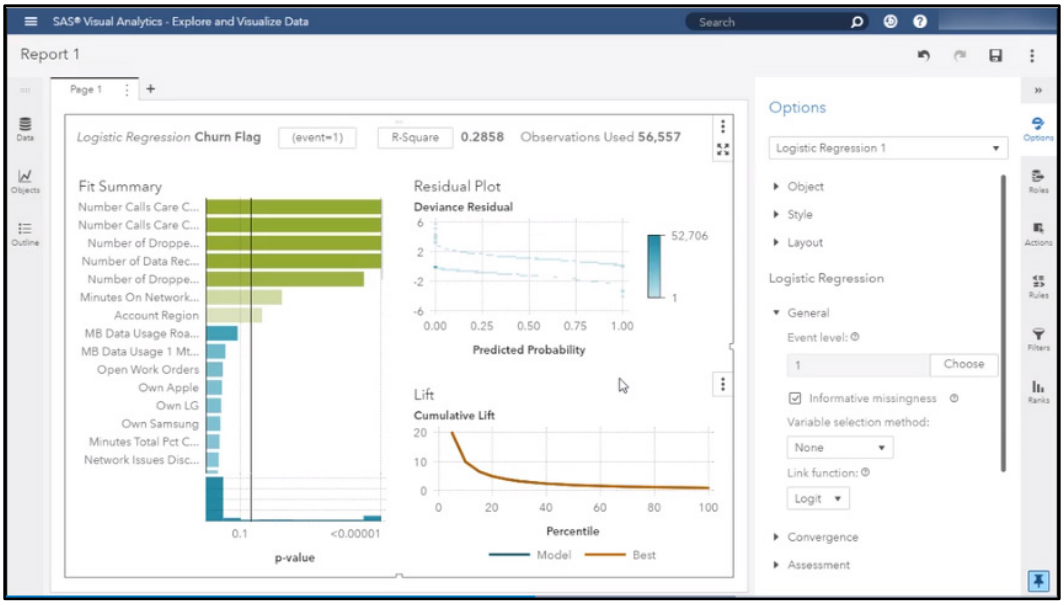


Missing Values

We see that a large portion of observations (41,785) are unused because of missing values. This Unused number is located in the upper left corner of the canvas. SAS Visual Statistics provides a convenient way to handle missing values in a regression model.

Choose Options from the far-right pane. Under the Logistic Regression Section, under General, click the check box to turn on Informative Missingness to automatically impute missing values and create the missing value flags, as shown in Figure 2.21.

Figure 2.21: Modified Logistic Regression Results with Imputed Missing Values



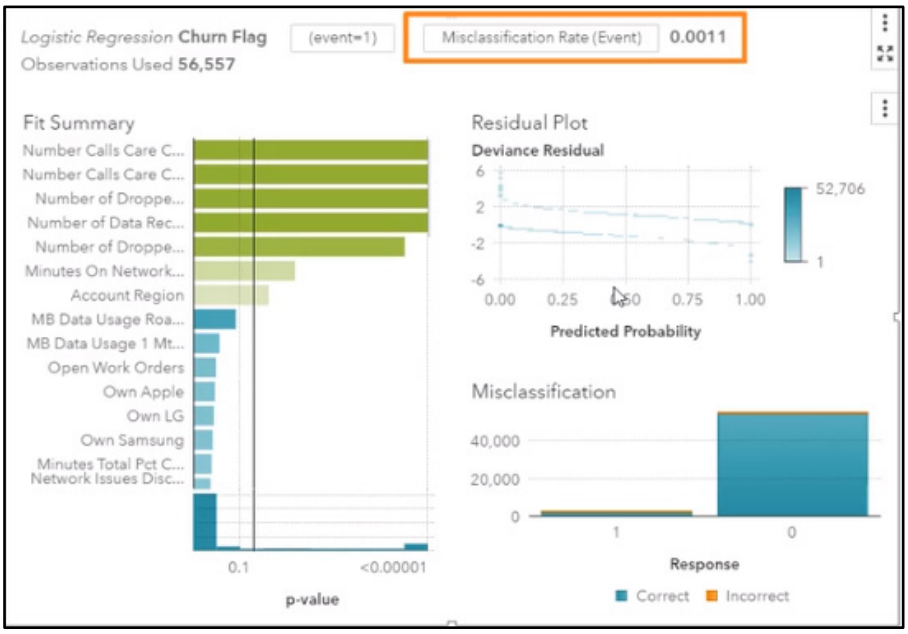
Model Assessment Visualizations

If you want to go further in your analysis, SAS Visual Statistics provides several model assessment visualizations, such as the Lift chart, the ROC curve, and the Misclassification chart. Right-click the Lift chart to view other options and change the chart type.

Misclassification Chart

The Misclassification chart enables you to evaluate the quality of your model, because you can see how many observations the model is correctly and incorrectly predicting, as shown in Figure 2.22. From the Misclassification Rate outlined in an orange box in Figure 2.22, we can conclude that we have a good model.

Figure 2.22: Misclassification Chart

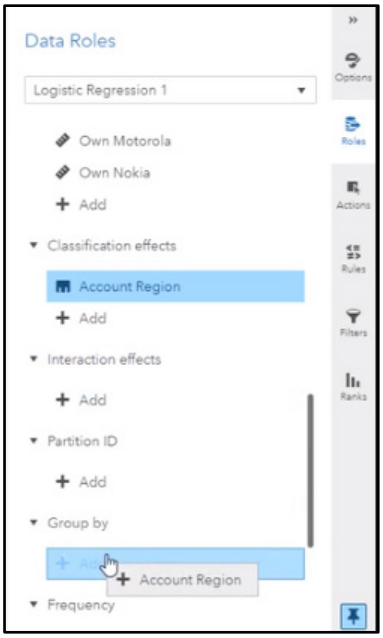


Group-by Role

Next, let's look at a powerful feature in SAS Visual Statistics: the ability to assign attributes to a group-by role. This means that, rather than building one regression model on the entire audience, you can build multiple regression models based on an attribute that you assign to this group-by role.

Let's look at how to do so in SAS Visual Statistics. In the Roles option in the far-right pane, move the variable named Account Region from the Classification effects list to the Group-by list. Select Account Region from the Classification effects category and drag it downward to the Group-by category, as shown in Figure 2.23.

Figure 2.23: Roles Pane



As you select it and assign it to the group-by role, SAS Visual Statistics very efficiently creates nine unique regression models for each particular region, as shown in the Fit Summary plot in Figure 2.24.

Figure 2.24: Modified Logistic Regression Results with Group-By Categories

So what is happening here in the lower-left plot? We have built a model for accounts from Great Lakes, accounts from Greater Texas, and so on. You can see how the variables contribute differently to each model in this heat map. For example, Number Calls Care Center is a significant attrition factor for only four of the customer regions, including Mid Atlantic, Pacific, South, and Great Lakes.

Actions

Now, suppose that we are happy with this analysis. How can we take action on this model? We have some options. Click on the drop-down button in the upper right corner of the canvas to see all actions, as shown in Figure 2.25.

Figure 2.25: Logistic Regression Actions

Export Model

First, we could export the score code. If you click the Export Model option, SAS Visual Statistics auto-generates the score code, and you can use that code to score a new data set in Base SAS or SAS Studio.

Derive Predicted Value

You can also derive predicted values from the model and use them directly in a Visual Analytics report.

Export Data

You can export model summary tables as Microsoft Excel files.

Resources

This chapter is based on the "SAS Visual Analytics on SAS Viya" videos in SAS® Viya® Enablement, a free course available from SAS Education.

You might find the following documentation helpful as you learn more about programming in SAS Visual Statistics:

- SAS Visual Analytics Documentation home
- SAS Visual Statistics Documentation home
- SAS® Visual Statistics in SAS® Viya®: Interactive Model Building course from SAS Education
- SAS Visual Analytics Communities

Chapter 3: SAS Visual Text Analytics

Introduction **61**
Batch Score a Topic Model **61**
Open a Project 61
View Data Source 62
View Pipeline 63
Download Score Code 64
Import and Load Scoring Data 64
Scoring Data 65
Create Custom Concepts **67**
View a Pipeline 67
View Custom Concepts 68
Create Custom Concepts 70
Explore Category Output Data Interactively **71**
View Category Output 72
Save Output Data 74
Explore in SAS Visual Analytics 74
Resources **79**

Introduction

SAS Visual Text Analytics is a web-based text analytics application that combines text mining, contextual extraction, categorization, and sentiment analysis.

SAS Visual Text Analytics analyzes large volumes of unstructured data, using predefined templates, machine learning methods, and natural language processing (NLP), to produce deeper insights from more data, faster than ever before. This software combines text mining, contextual extraction, categorization, sentiment analysis, and search within a modern and flexible framework. An end-to-end visual pipeline makes it easy to prepare data, visually explore topics in SAS Visual Analytics, extract entities and facts, analyze sentiment, build text models, and deploy models within existing systems or processes.

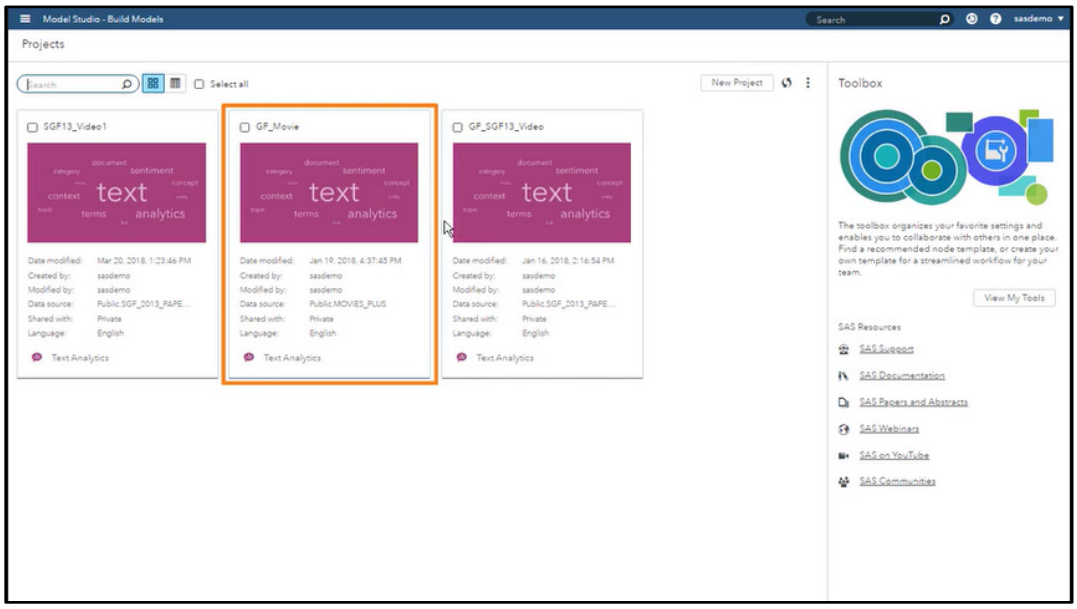
Batch Score a Topic Model

In this section, we will learn how to use the SAS Visual Text Analytics topics model to score external text data in SAS Model Studio and SAS Studio. This functionality enables you to discover the themes or topics that are found in your document collection.

Open a Project

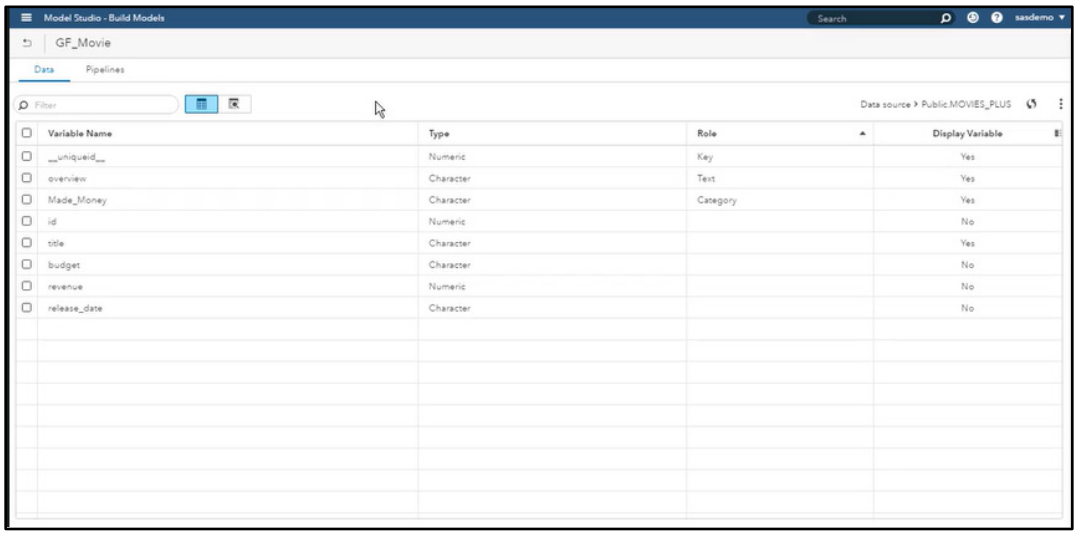
Let's start in SAS Model Studio. In this example, we will begin by opening an existing SAS Model Studio project named GF_Movie, as shown in Figure 3.1.

Figure 3.1: SAS Model Studio Home Page



By default, the Data Attributes window appears, as shown in Figure 3.2. The data source is shown in the upper right corner. In this example, the data source is a Cloud Analytics Services (CAS) data table called MOVIES_PLUS. The variables of interest are uniqueid (system-generated ID variable), overview (movie review text variable), and Made_Money (binary categorical variable).

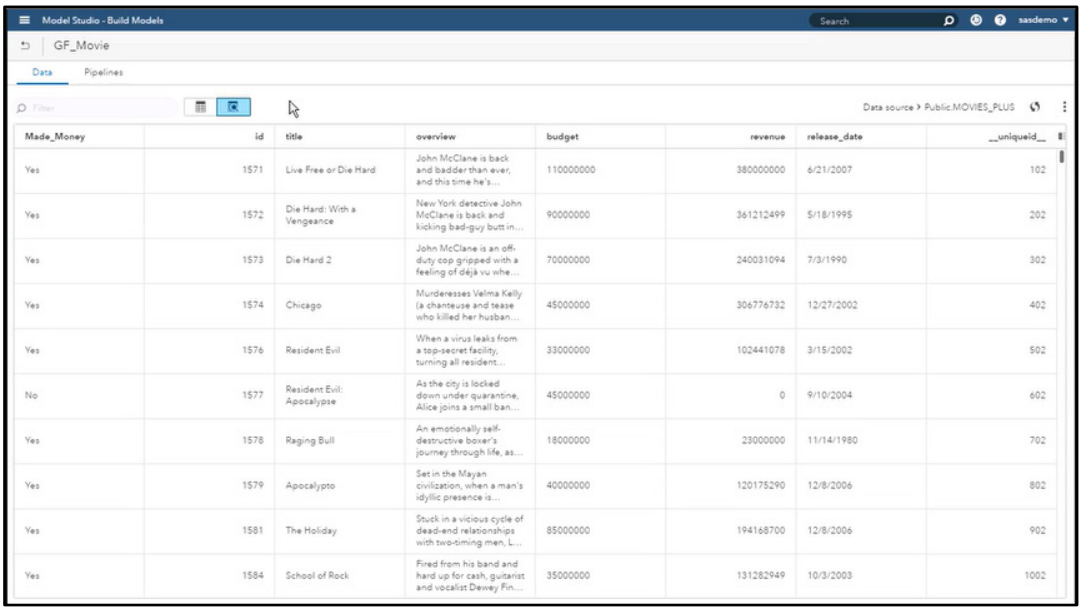
Figure 3.2: Data Attributes Window



View Data Source

Click the Data button, which is the second button on the right, next to the search filter box, as shown in Figure 3.3. Notice that the first few rows of the MOVIES_PLUS data appear.

Figure 3.3: Sample Data Rows



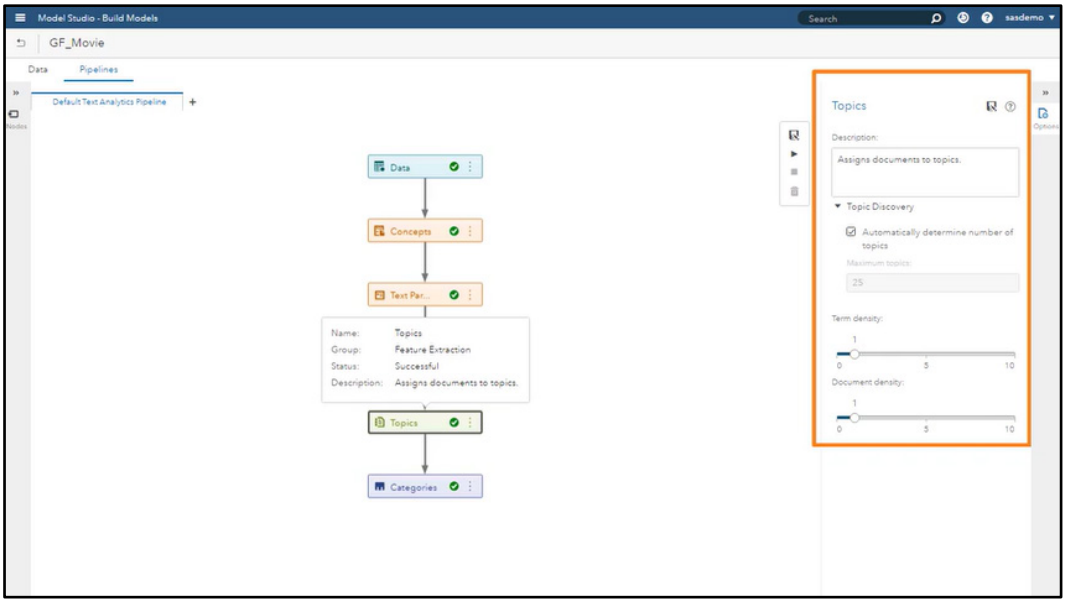
Made_Money	id	title	overview	budget	revenue	release_date	__uniqueid__
Yes	1571	Live Free or Die Hard	John McClane is back and badder than ever, and this time he's...	110000000	380000000	6/21/2007	102
Yes	1572	Die Hard: With a Vengeance	New York detective John McClane is back and kicking bad-guy butt in...	90000000	361212499	5/18/1995	202
Yes	1573	Die Hard 2	John McClane is an off-duty cop gripped with a feeling of déjà vu whe...	70000000	240031094	7/3/1990	302
Yes	1574	Chicago	Murderesses Velma Kelly (a chanteuse and tease who killed her husban...	45000000	306776732	12/27/2002	402
Yes	1576	Resident Evil	When a virus leaks from a top-secret facility, turning all resident...	33000000	102441078	3/15/2002	502
No	1577	Resident Evil: Apocalypse	As the city is locked down under quarantine, Alice joins a small ban...	45000000	0	9/10/2004	602
Yes	1578	Raging Bull	An emotionally self-destructive boxer's journey through life, as...	18000000	23000000	11/14/1980	702
Yes	1579	Apocalypto	Set in the Mayan civilization, when a man's idyllic presence is...	40000000	120175290	12/8/2006	802
Yes	1581	The Holiday	Stuck in a vicious cycle of dead-end relationships with two-timing men, L...	85000000	194168700	12/8/2006	902
Yes	1584	School of Rock	Fired from his band and hard up for cash, guitarist and vocalist Dewey Fin...	35000000	131282949	10/3/2003	1002

View Pipeline

To view the pipeline, select the Pipeline tab from the top left of the window, and then select Default Text Analytics Pipeline from the tab that appears beneath it. Pipelines are structured flows of analytic actions. These analytic actions are represented as individual nodes in a pipeline.

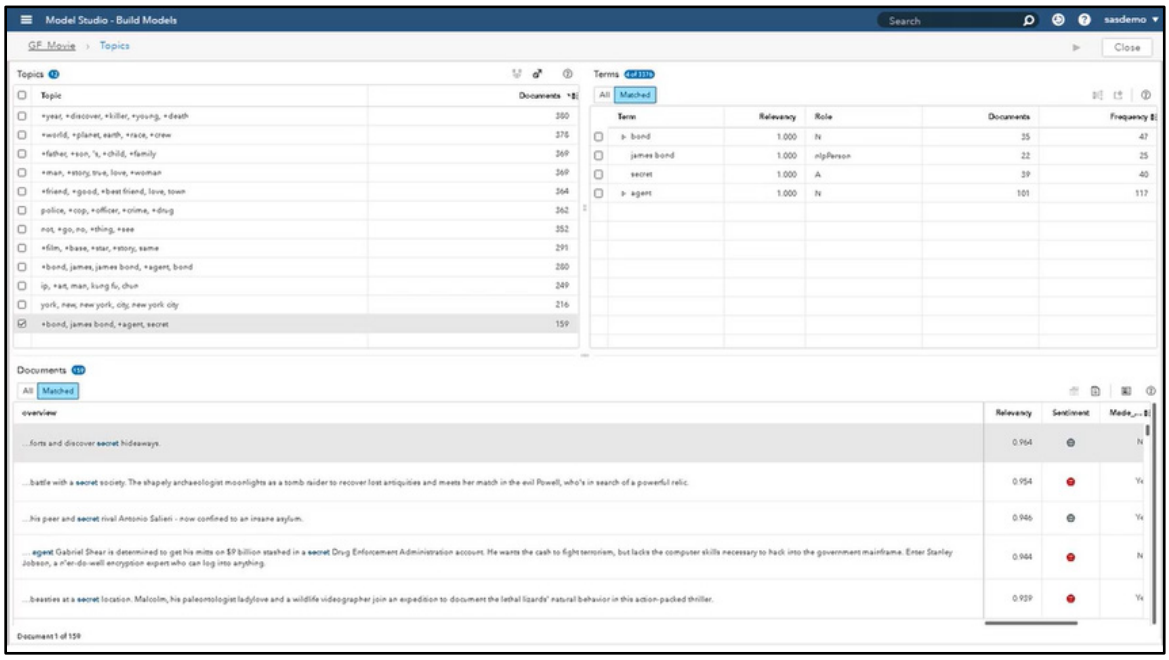
One of the nodes in the Default Text Analytics Pipeline is Topics. This node groups similar documents into related themes, or topics, by using a machine learning algorithm. Click the Topics node. Notice that the default settings for the node appear to the right, as shown in Figure 3.4.

Figure 3.4: Pipeline Tab



To view and interact with system-generated topics, right-click the Topics node and select Open from the drop-down menu. Doing so opens the interactive Topics window. The interactive Topics window displays all the topics that SAS Visual Text Analytics identified, as shown in the upper-left quadrant of Figure 3.5. The default name of a topic is the top five most relevant terms for that topic.

Figure 3.5: Topics Window



Examine the system-generated topics. The ability to create user-defined topics from user-specified terms is a new feature in SAS Visual Text Analytics. In addition, you can merge two or more topics, or split a topic if its definition is too broad. For each selected topic, you can also view the matched terms in the upper-right quadrant of the window. The matched terms within the matched document, the relevancy score, and the sentiment classification are all shown in the bottom half of the window.

Close the Topics output window by clicking the Close button in the upper-right corner. You are returned to the pipeline.

Download Score Code

Let's download the score code from the Topics node. Right-click the Topics node in the pipeline, and select Download score code from the drop-down menu. The score code is downloaded as a zip file named TopicScoreCode(2) and is stored in the default download folder.

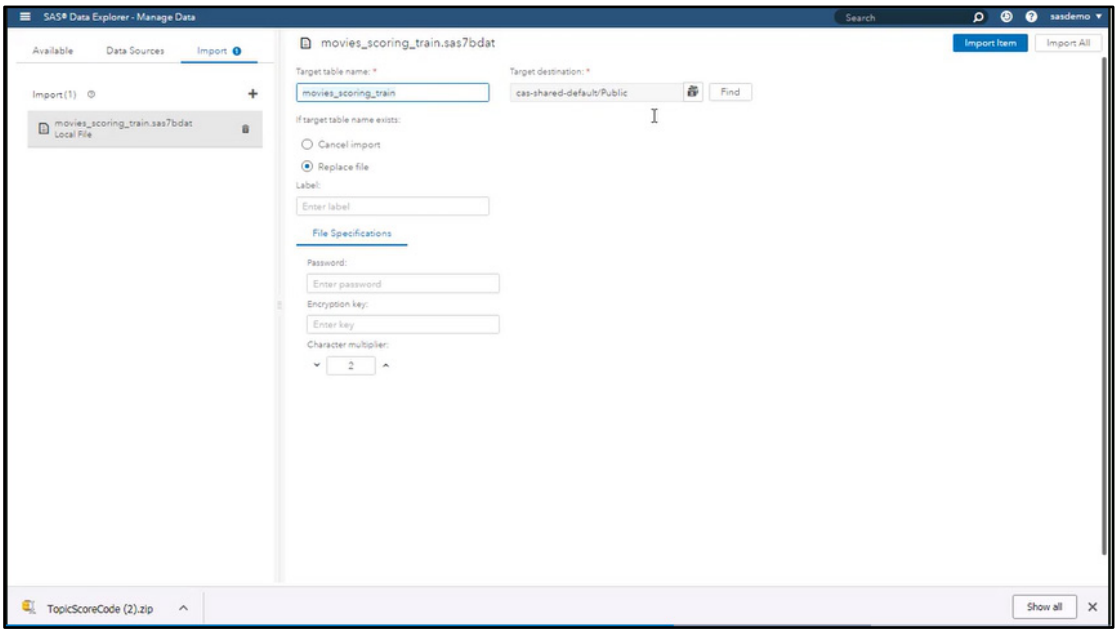
Import and Load Scoring Data

The first step in scoring an external text data file in SAS Studio is to import and load the scoring data, moving_scoring_train, into the caslib.

Open SAS Data Explorer by clicking the three horizontal bars in the upper-left corner of the screen, and then select Manage Data. Select the Import tab from the panel on the left, and then select Local File. Navigate to the data file and click Open.

Specify the target table name movies_scoring_train, as shown in Figure 3.6. For the target destination, select the default library: cas.shared.default/Public. Click Import Item.

Figure 3.6: SAS Data Explorer Import Data Tab

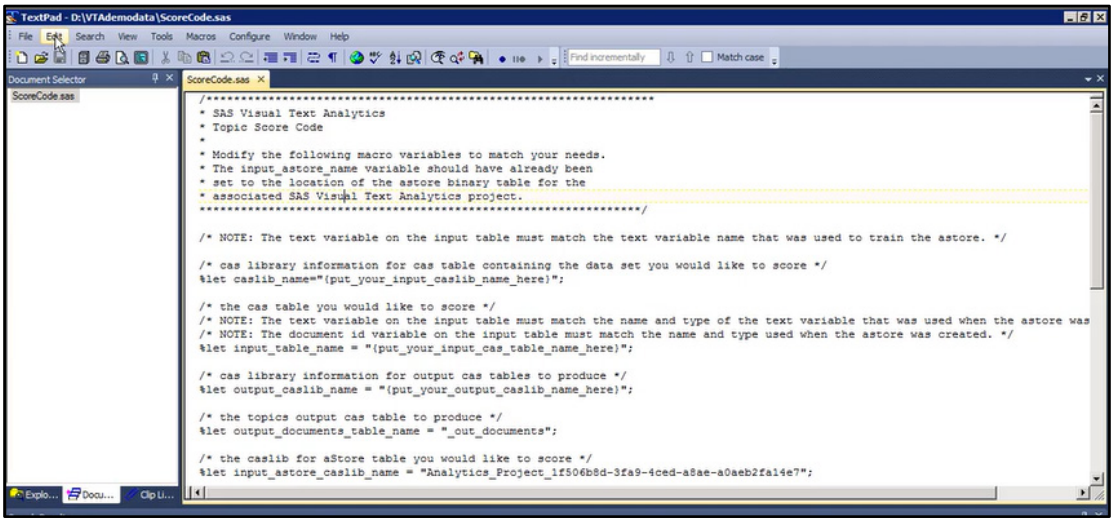


A note with a green check mark appears, indicating that the local data was successfully imported as a CAS table. Now we are ready to score the external data in SAS Studio.

Scoring Data

Unzip the previously saved score code that you downloaded in a previous step as TopicScoreCode(2).zip. Open the file scorecode.sas in a text editor such as TextPad or Notepad, and copy the contents of scorecode.sas to the clipboard, as shown in Figure 3.7.

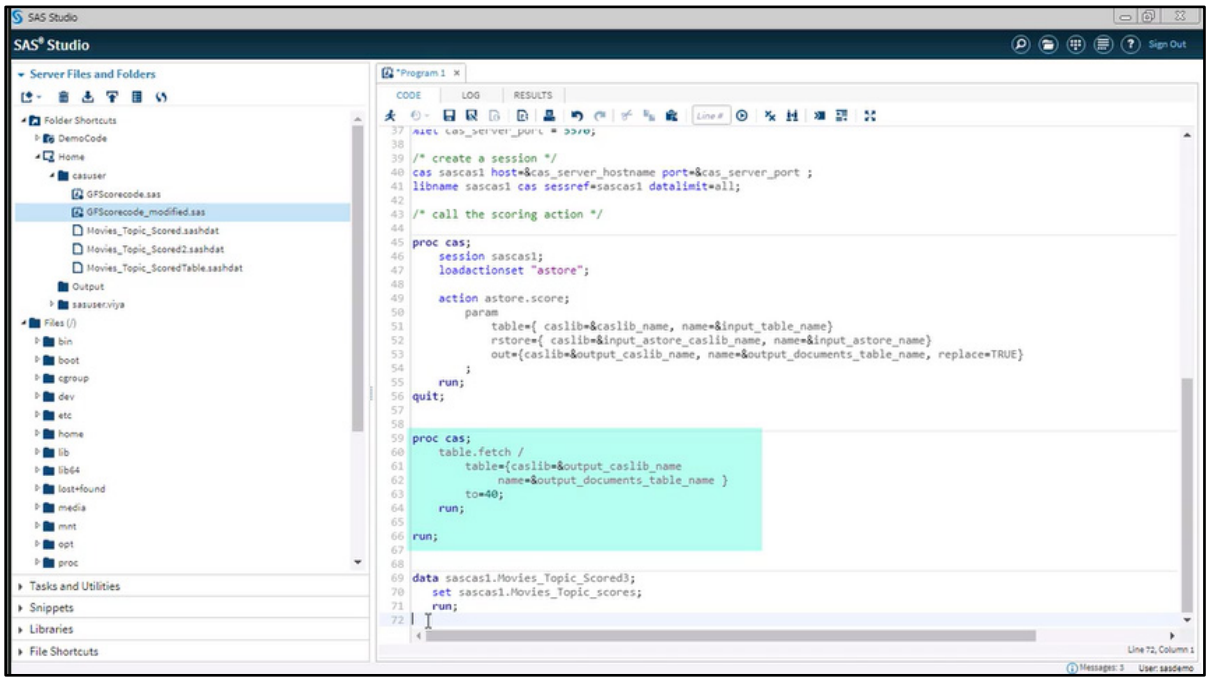
Figure 3.7: Text Editor Window



Select all text and hit CTRL+C or choose copy.

Start SAS Studio. Open the code window, and paste the contents of ScoreCode.sas from the clipboard to the window. To view a partial list of the scored output table, let's append the score code with a PROC CAS step to view 40 rows of the output table, as shown in Figure 3.8.

Figure 3.8: SAS Studio Macro Code

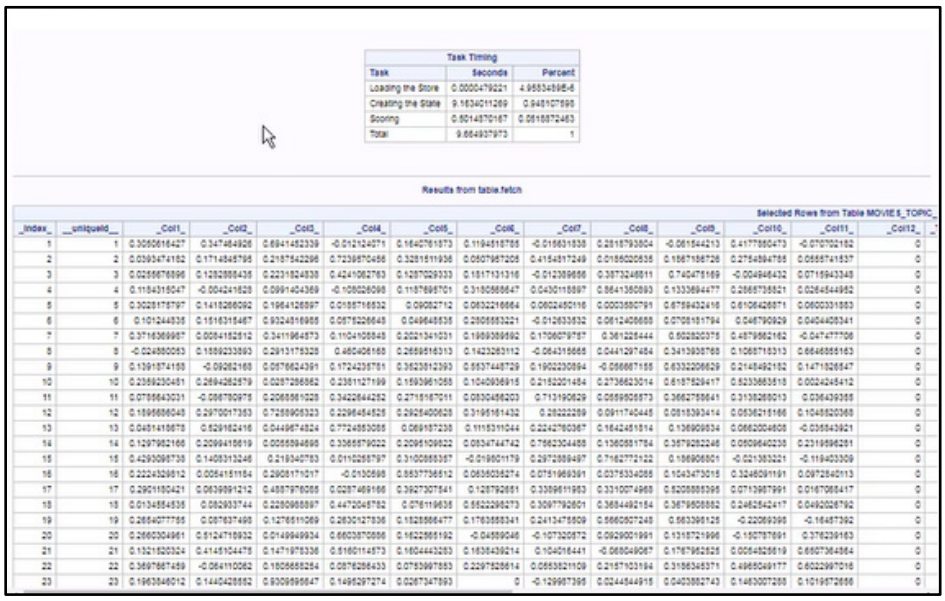


Before we can submit the SAS code, we need to specify inputs for some of the macro input fields. In the Score Code macro, look for snippets with {put_your_input_caslib_name_here} and replace them as follows:

```
%let caslib_name = "PUBLIC";

%let input_table_name = "Movies_Scoring_Train";

%let output_caslib_name = "PUBLIC";

%let output_documents_table_name = "Movies_Topics_scores";
```

Submit the modified code by clicking the running man icon. Check the log to verify that no errors are reported while scoring. Now let's view the results in Figure 3.9.

Figure 3.9: Results of ScoreCode.sas



The output data contain document scores that correspond to each topic. The output data also include the results of binary topic classification scores for each document in the external data.

Create Custom Concepts

In this section, we will learn how to create custom concept rules automatically by using the Concepts nodes in SAS Visual Text Analytics. The Concepts node enables you to identify and extract entities in a document collection by doing one of the following:

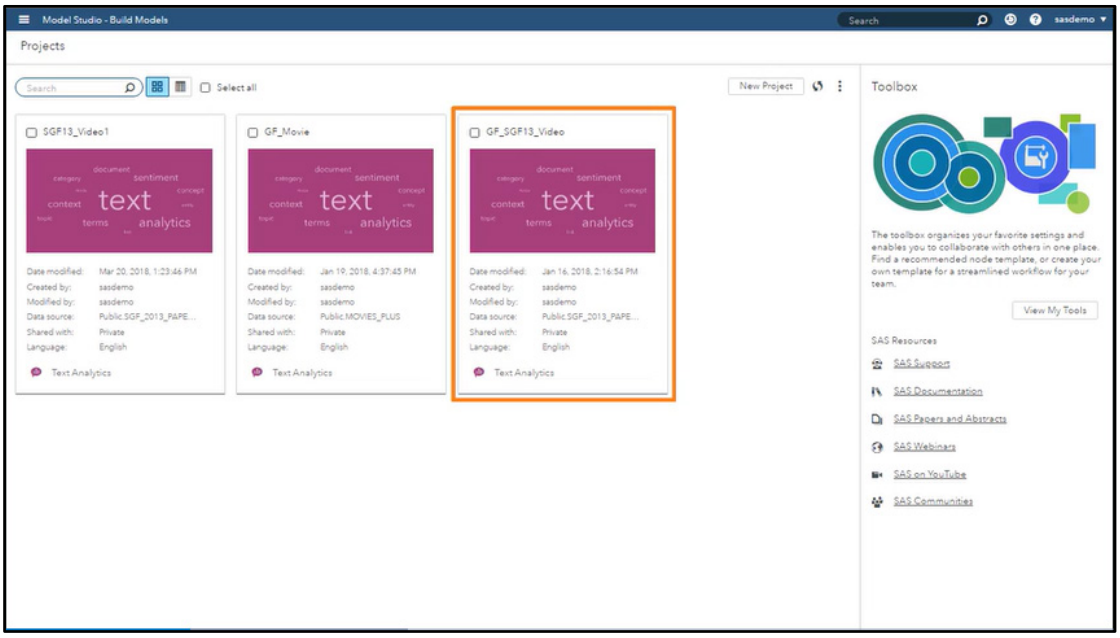
- Using up to nine predefined concepts
- Creating additional custom concept rules
- Using a combination of both predefined and custom rules

In addition, when the Concepts node follows the Text Parsing node in the analytic pipeline, the kept textual elements are available in the concept window. This new feature enables analysts to group similar textual elements in the Concepts node window and then use selected textual elements to develop automatic concept rules.

View a Pipeline

We will start in SAS Model Studio and use the existing SAS Model Studio project named GF_SGF13_Video, outlined in an orange box in Figure 3.10.

Figure 3.10: SAS Model Studio Home Page

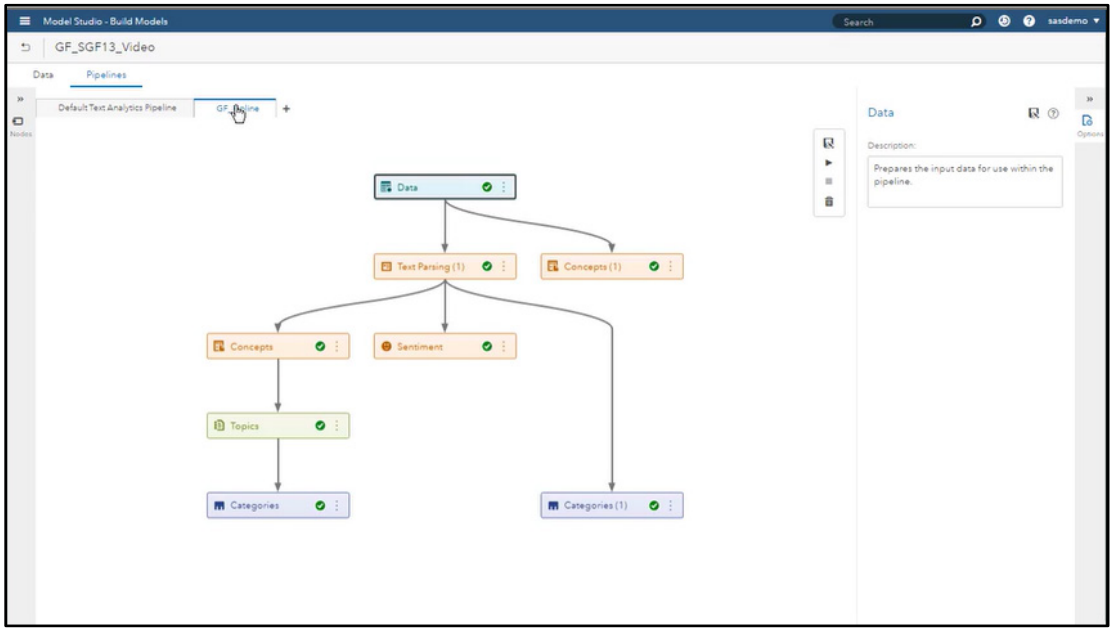


Select the project by clicking on it to open it. Notice that the data attributes window appears. Click the Pipelines tab, and then select the previously created pipeline named GF_Pipeline. By default, SAS Visual Text Analytics provides six text analysis pipeline nodes, as shown in Figure 3.11. These nodes are as follows:

- Data
- Text parsing
- Concepts
- Sentiment
- Topics
- Categories

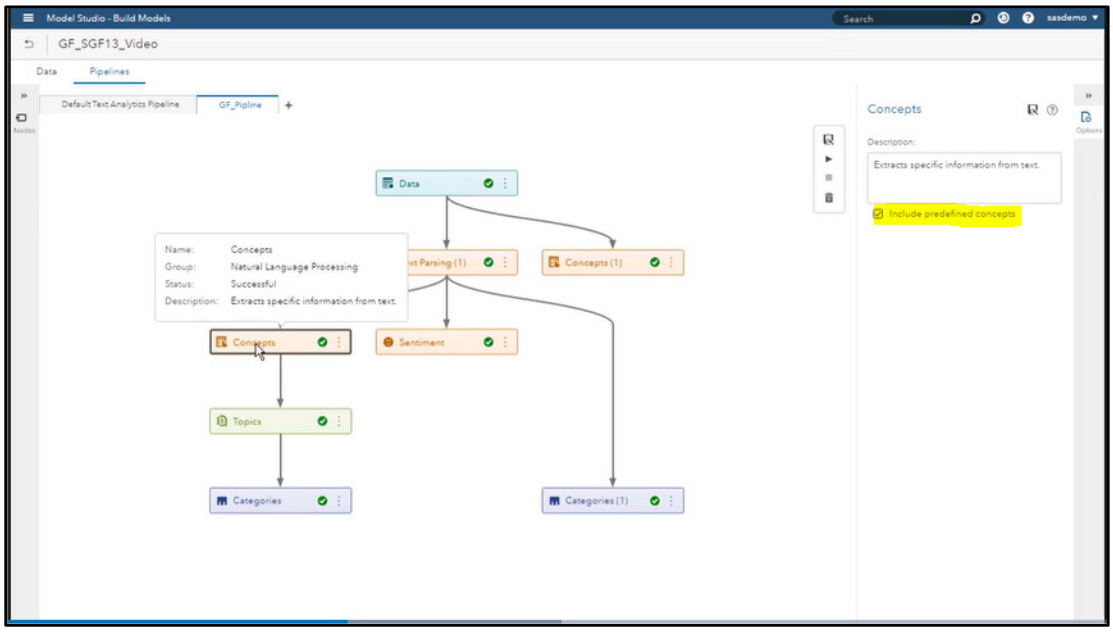
The nodes are arranged in a specific sequence that users can control in accordance with their analytic goal. Each node is designed to solve a specific text analytics problem.

Figure 3.11: Custom Pipeline



The custom pipeline in Figure 3.11 includes a Concepts node, which is placed after the Text Parsing node. If you click a node, you can examine the default settings. For example, if you click the Concepts node, as shown in Figure 3.12, you will notice that in the Concepts pane on the right, the only option you can specify for the Concepts node is whether to include predefined concepts. All nine predefined concepts are included in the analysis and are standardized across all languages.

Figure 3.12: Concepts Node in Custom Pipeline

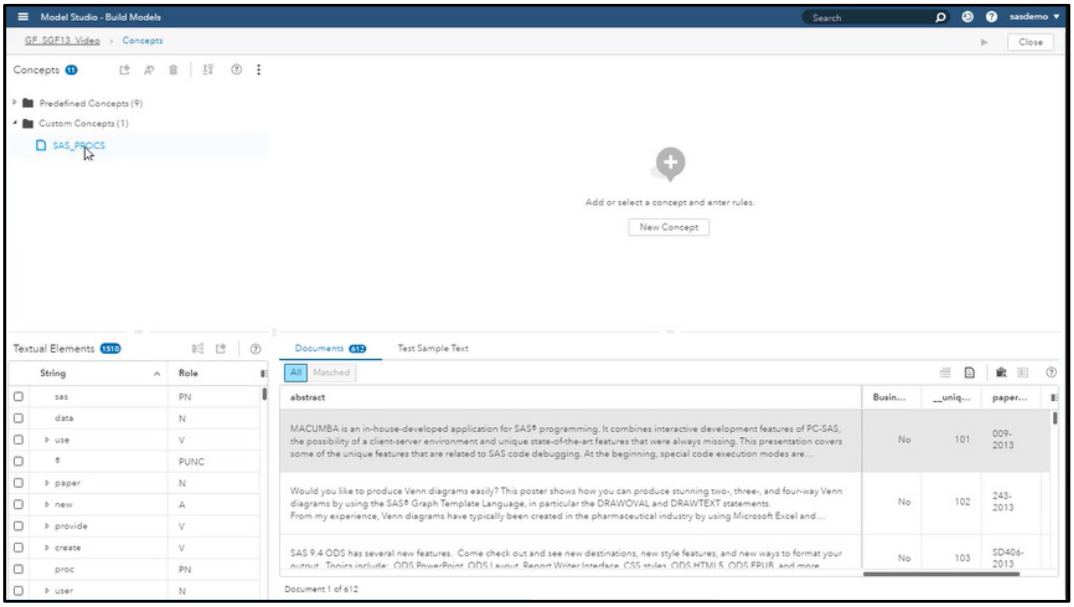


View Custom Concepts

Right-click the Concepts node, and select Open to view document matches on predefined concepts and to create custom concepts that enable you to extract domain-specific information. In the upper-left of the Concepts node

results window, the nine predefined concepts are displayed, as shown in Figure 3.13. These are concepts whose rules are already written.

Figure 3.13: Concepts Window

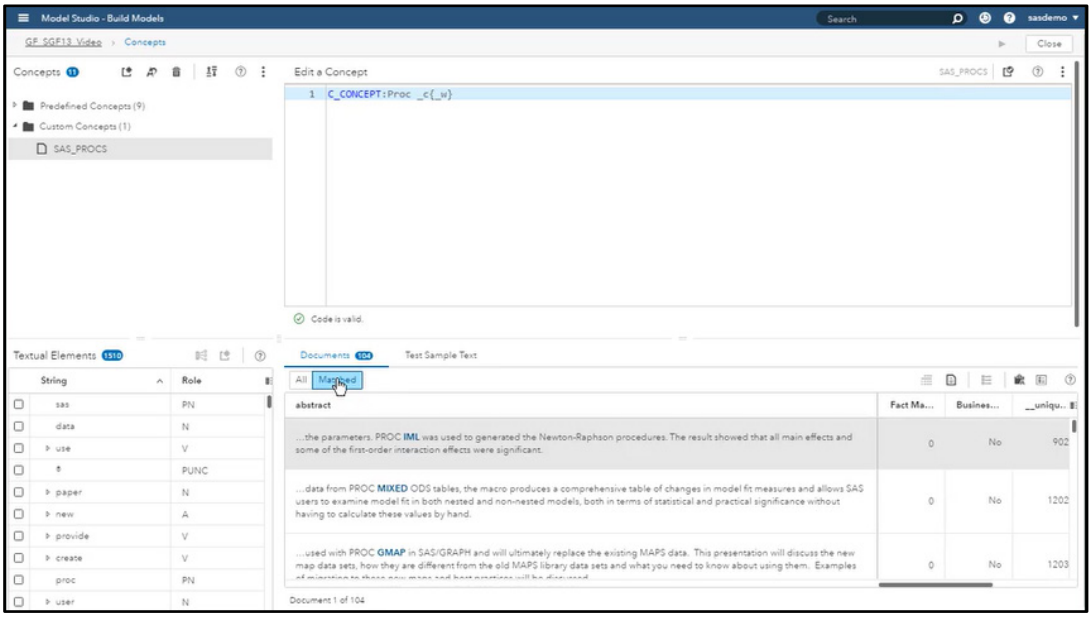


You cannot rename predefined concepts or view and edit their base rule definitions. However, you can add rules by using the Edit window to modify or extend predefined concept behavior. You can create custom concepts by using custom rules. In addition, you can prioritize which matches are returned when overlapping matches occur.

Let's explore a custom concept that has already been made in the system. In the Custom Concepts category in the upper right, select SAS_PROCS. When you are naming a custom concept, it is important to follow the custom concept naming guidelines. Valid custom concept names can include numbers, letters, and matched underscores on both sides. It is a good idea to avoid using common terms that could be matched in your document.

The rule definition associated with SAS_PROCS is a C_CONCEPT rule, as shown in the edit window of Figure 3.14. This custom concept extracts all SAS PROCs referenced in the abstracts collection.

Figure 3.14: Edit Concept



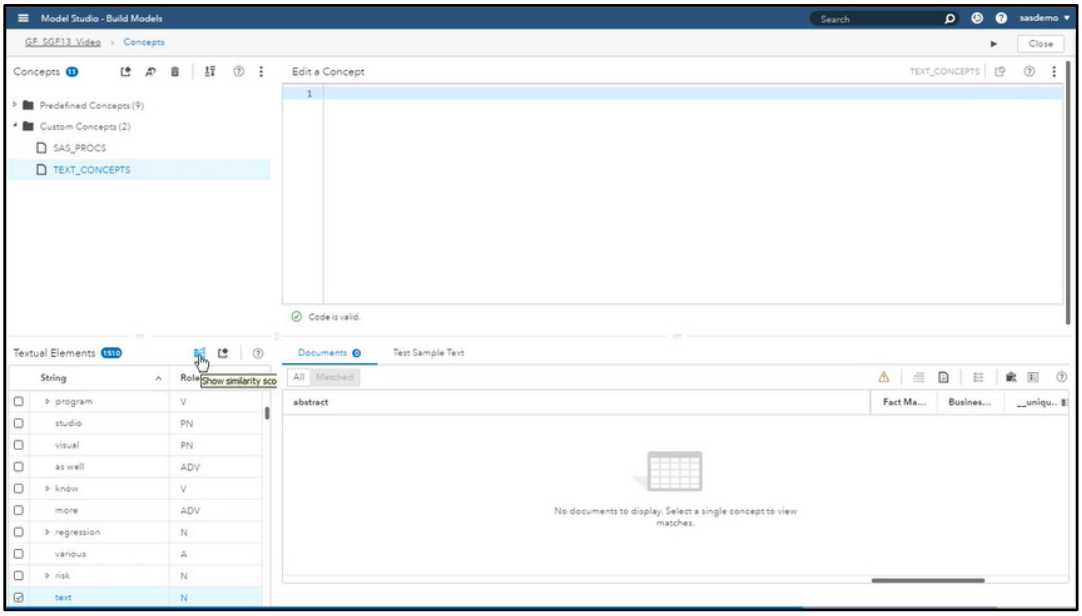
Notice the Matched Documents in the Documents window below the edit window. All the SAS PROCs included in the abstracts are highlighted in the documents. If a matched item, matched fact, or overlapping match is discovered, the match is differentiated by certain visual cues.

Create Custom Concepts

Now let's create a custom concept rule automatically by including specific textual elements in the document collection. Select the Custom Concepts folder in the upper left, and then select the New Concept button in the center of the screen. Name the custom concept TEXT CONCEPTS. We are now ready to create a custom concept automatically by selecting the main textual element.

In the Textual Elements window in the bottom left, scroll and select the string "text," as shown in Figure 3.15. Click the Similarity icon to generate the similarity scores related to the string text.

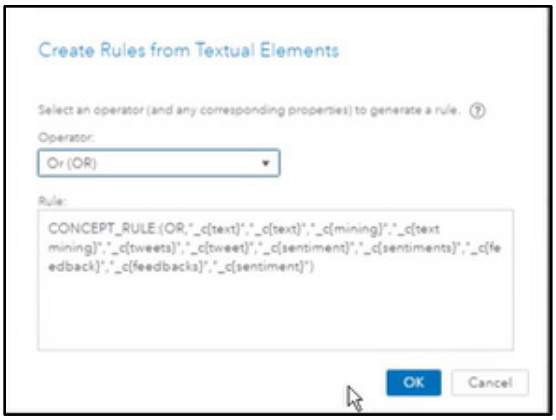
Figure 3.15: Similarity Icon



The higher the similarity score, the more likely the string is to appear in the same context as the selected string. In this example, we will select all strings related to text. Then, click the Create Rules from Selected Elements icon.

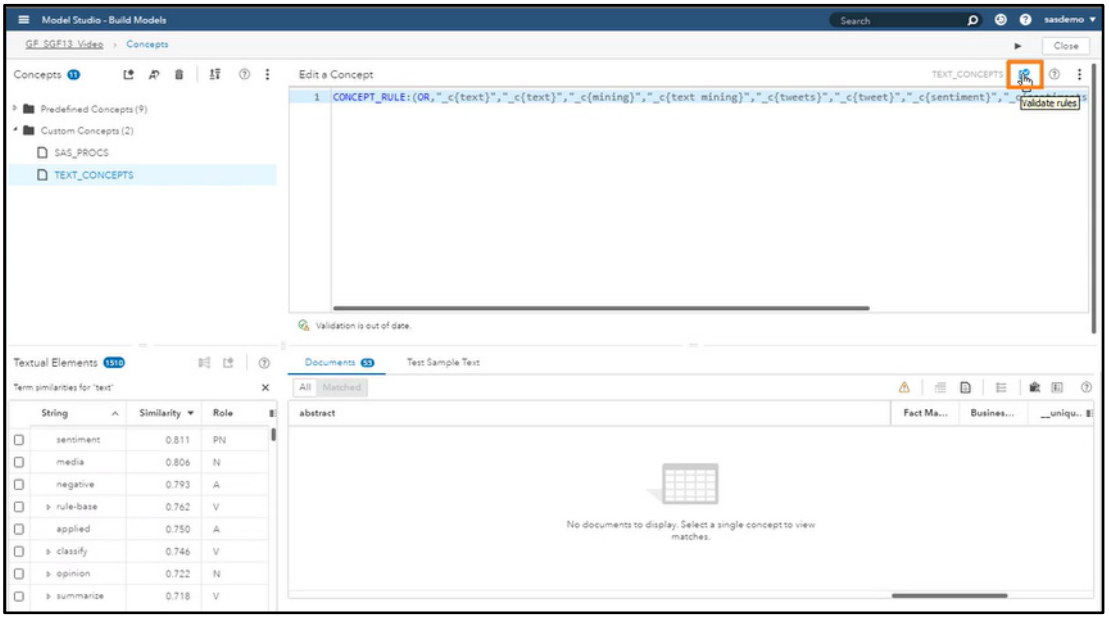
The SAS Visual Text Analytics Concepts node automatically generates a CONCEPT_RULE, using all the selected strings, and displays the rule syntax in the Create Rules from Textual Elements window, as shown in Figure 3.16. If necessary, you can select an alternate operator from the drop-down list or modify the system-generated rule definition.

Figure 3.16: Create Rules from Textual Elements Window



When you are satisfied with the rule definition, click OK. You return to the main Concepts screen, as shown in Figure 3.17. Click the Validate icon rule to validate the new rule syntax.

Figure 3.17: Validate Rule Syntax



If the code can be validated, then the "Validation is out of date" message beneath the edit window will change to "Code is valid."

In the Documents window in the bottom middle of the screen, notice that the Matched tab shows the extracted entities for each document. Now you have seen how the automated concept-rule-writing feature using textual elements in SAS Visual Text Analytics can help simplify concept extraction.

Explore Category Output Data Interactively

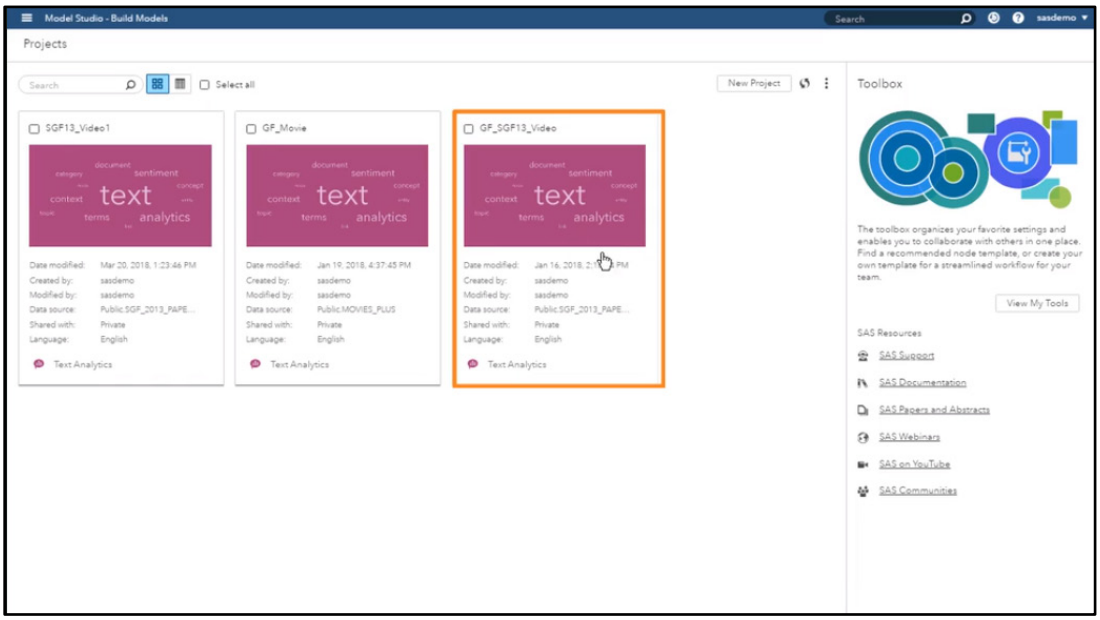
In this section, we will learn how to interactively explore category model output data in SAS Visual Text Analytics. A category identifies a group of documents that share a common characteristic. In SAS Visual Text Analytics, categories are generated by the following actions:

- Promoting a topic to a category
- Specifying a category variable while creating a new project
- Creating a new category in the Categories nodes

After generating initial category rules, you can edit the rules for optimal categorization. You can then save and explore the output data generated by the Categories node.

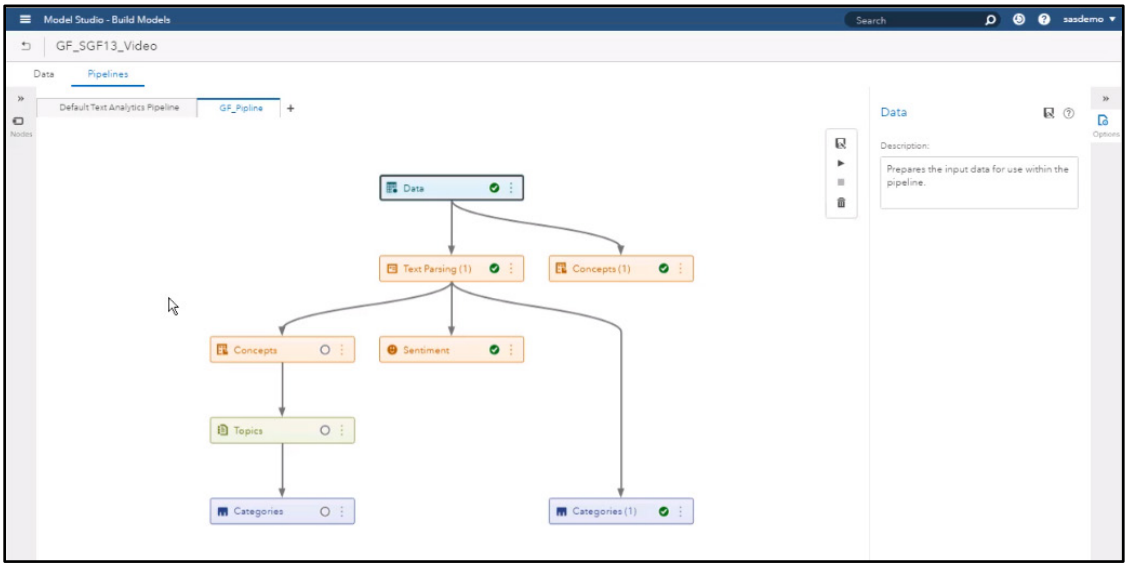
We will demonstrate this in SAS Model Studio with an existing SAS Model Studio project. Select the project named GF_SGF13_Video, as shown in Figure 3.18.

Figure 3.18: SAS Model Studio Home Page



By default, the data attributes window opens. Click the Pipelines tab. Then, select the previously created pipeline named GF_Pipeline, as shown in Figure 3.19. SAS Visual Text Analytics provides six text analysis pipeline nodes, arranged in a sequence that you can control. Each of the nodes is designed to solve a specific text analytics problem.

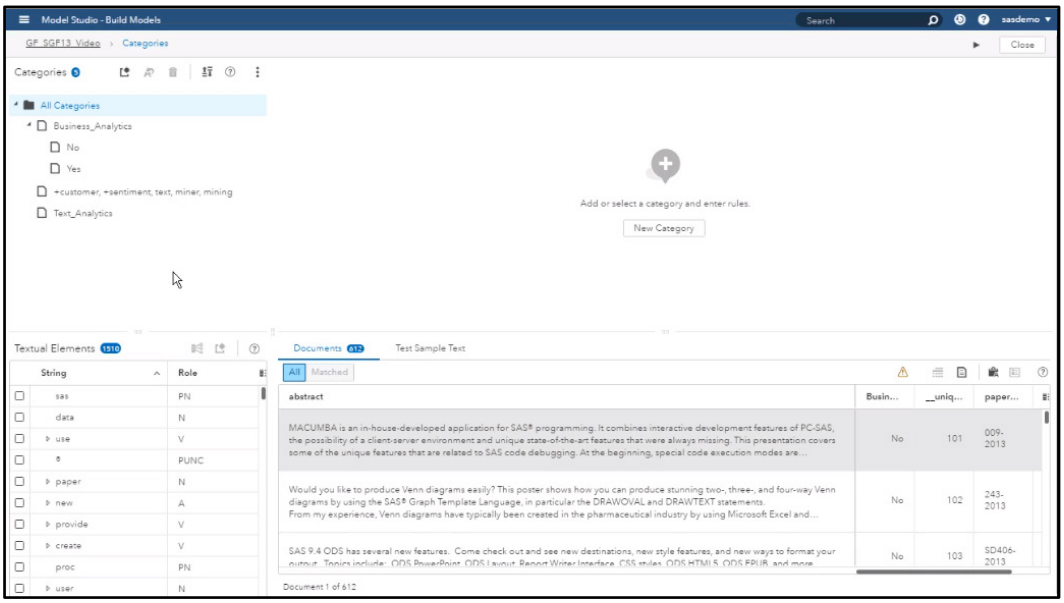
Figure 3.19: GF_Pipeline



View Category Output

Click the Categories node at the bottom left of the pipeline. Notice that the default settings appear on the right. Let's view the category output. Right-click the Categories node and select Open from the drop-down menu. This will open the Categories window, as shown in Figure 3.20.

Figure 3.20: Category Window

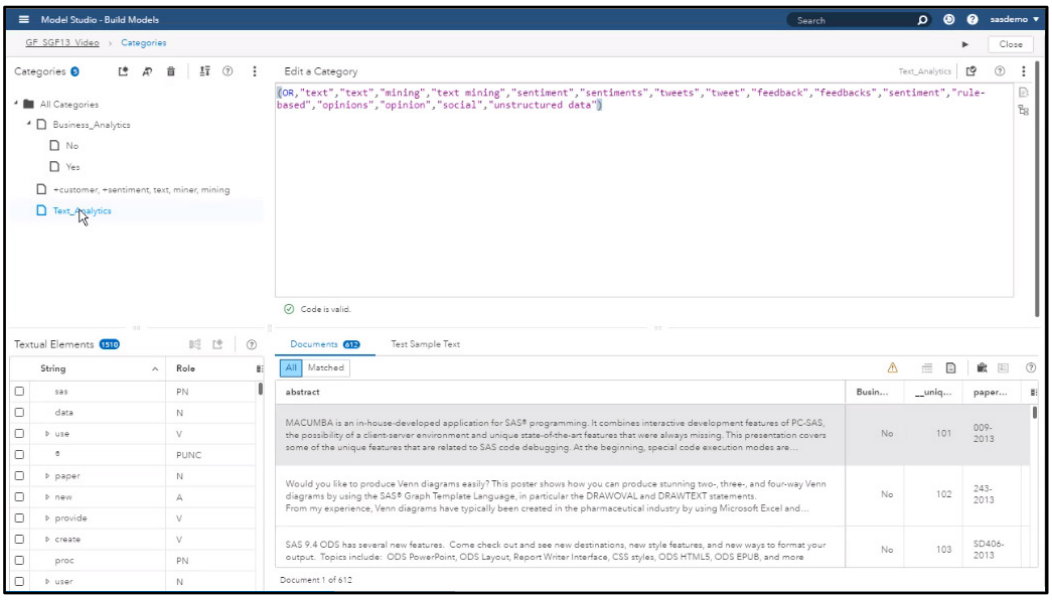


In the Categories window, notice that All Categories in the upper-left pane contains four subcategories:

- **Business_Analytics No**. This is a sub-subcategory within Business_Analytics. This sub-subcategory was manually classified in the training data before the project was created.
- **Business_Analytics Yes**. This is a sub-subcategory within Business_Analytics. This sub-subcategory was manually classified in the training data before the project was created.
- **+ customer, + sentiment, text, miner, mining**. This is a discovered topic that was promoted to a category in the previous Topic node.
- **Text_Analytics**. This is a new subcategory that was created by selecting text elements in the category results window. The Categories node created the associated rule from the selected textual elements in a previous run.

In the Edit a Category pane, notice the rules that were generated for the subcategories when you click on each subcategory. In addition, you can examine the relevancy scores and the matching documents for each of the subcategories, as shown in Figure 3.21.

Figure 3.21: Edit a Category



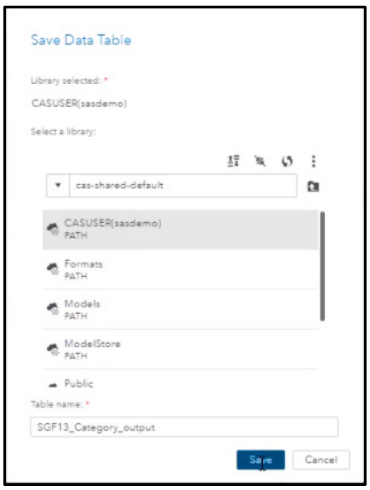
Note that the sentiment for each document is not displayed here, because a Sentiment Analysis node was not included in the pipeline preceding the Categories node. Click Close in the upper-right corner to return to the pipeline.

Save Output Data

Run the pipeline by clicking the Run shortcut button (▶) from the menu in the upper right.

Let's save the output data table so that we can explore it in SAS Visual Analytics. Right-click the Categories node and select Save data table from the drop-down menu. Name the output table "SGF13_Category_Output," and save it to the default caslib, CASUSER, as shown in Figure 3.22. Click Save.

Figure 3.22: Save Data Table Window

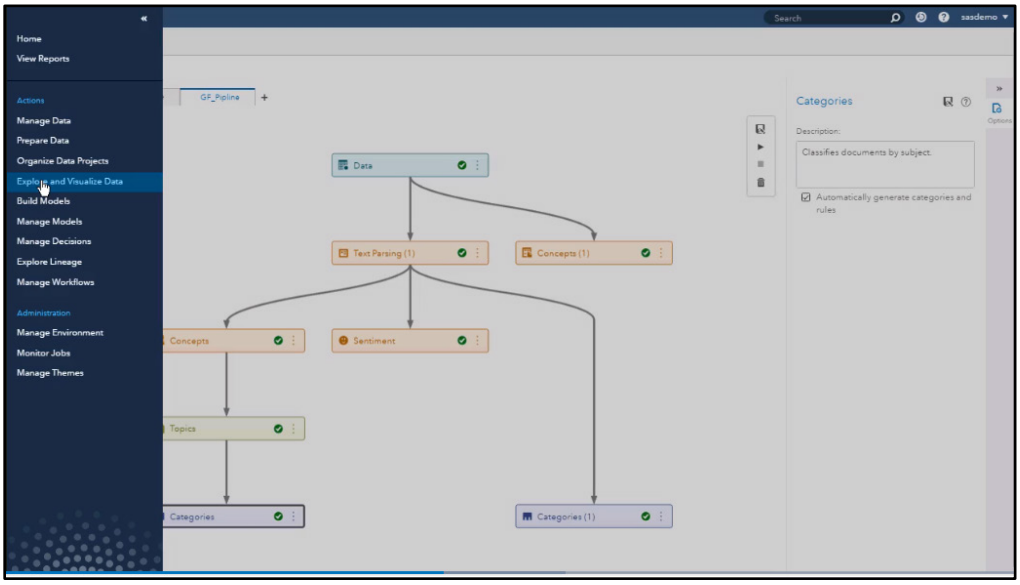


Now that we have saved the output data table, we are ready to explore and visualize it in SAS Visual Analytics.

Explore in SAS Visual Analytics

Click the three horizontal bars in the top left corner, which is the Show applications menu. Select Explore and Visualize Data as shown in Figure 3.23.

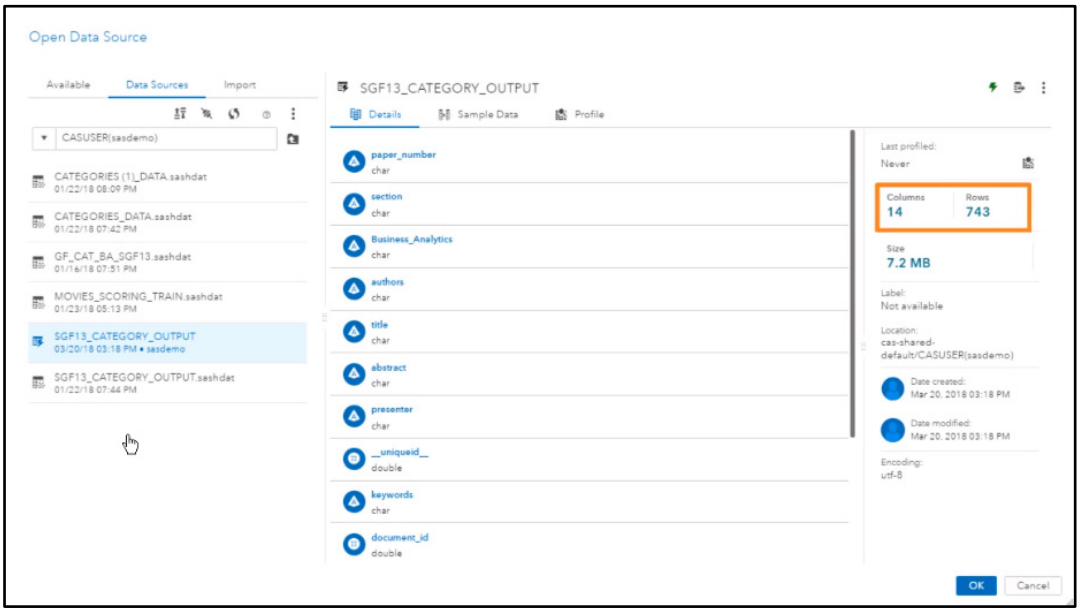
Figure 3.23: Applications Menu



The SAS Visual Analytics Explore and Visualize Data window appears. The first step in data visualization is choosing the data to explore. Click Data in the panel on the left side of the window. Select Add Data Source from the first drop-down menu. Click the Data Sources tab in the window that appears. Under CAS-shared-default, select CASUSER. In the list of available data tables, select SGF13_Category_output, which we created in the previous step. Right-click the data table name, and select Load from the drop-down menu.

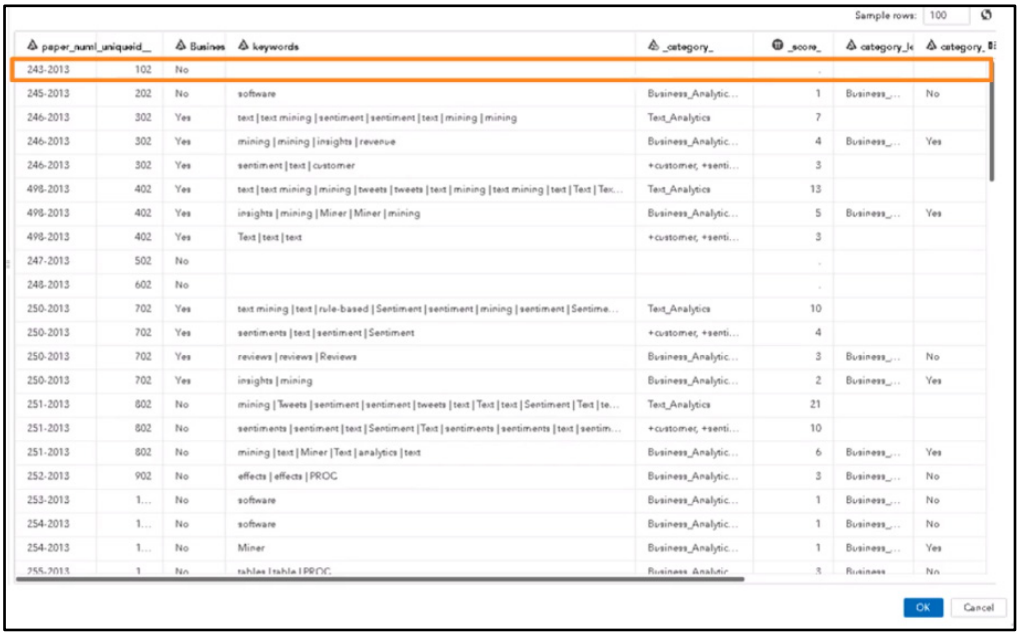
The details panel appears by default, as shown in Figure 3.24. Notice that the data contain 14 columns and 743 rows. Because a single document can be classified by more than one category, the number of rows increased from 612 to 743, and the total number of documents increased from 612 to 743.

Figure 3.24: Visual Analytics Data Sources Tab



Let's explore a sample of this data. Click the Sample Data tab in the middle window. By default, the first 100 rows are displayed, as shown in Figure 3.25. You can change the default value if you want to view a different number of rows. Notice that the number of category assignments varies. For example, as shown in the highlighted row in Figure 3.25, Document 102 is classified into the *missing* subcategory (that is, not assigned to any one of the four subcategories that we saw earlier).

Figure 3.25: Sample Data

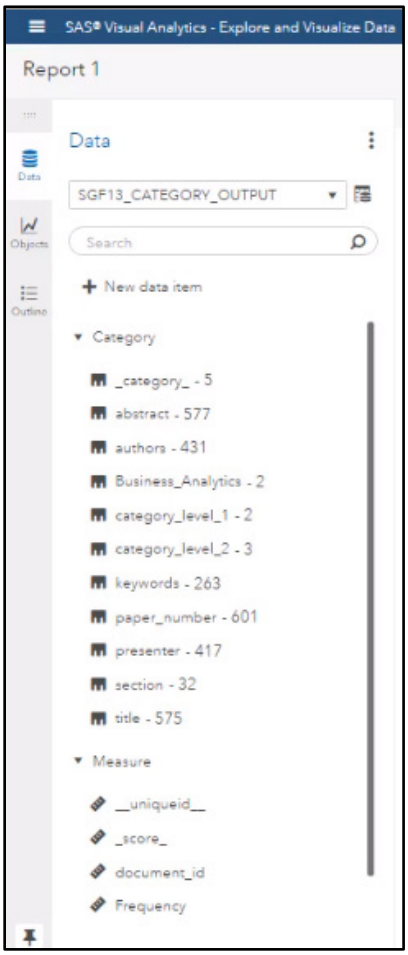


Meanwhile, in the next row, Document 202 is classified into the Business_Analytics No subcategory. Document 302 is classified into three different subcategories, as you can see in the next three rows.

Click OK to open the Explore and Visualize Data window. In addition to the original variables shown in the Data panel in Figure 3.26, several new variables were created by the Categories node:

- **_category_**. This is the assigned subcategory level.
- **Keywords**. This is the terms responsible for grouping that specific document into a specific category.
- **_score_**. This is the category membership scores, which indicate the number of keywords present in the document.

Figure 3.26: Visual Analytics Data Panel

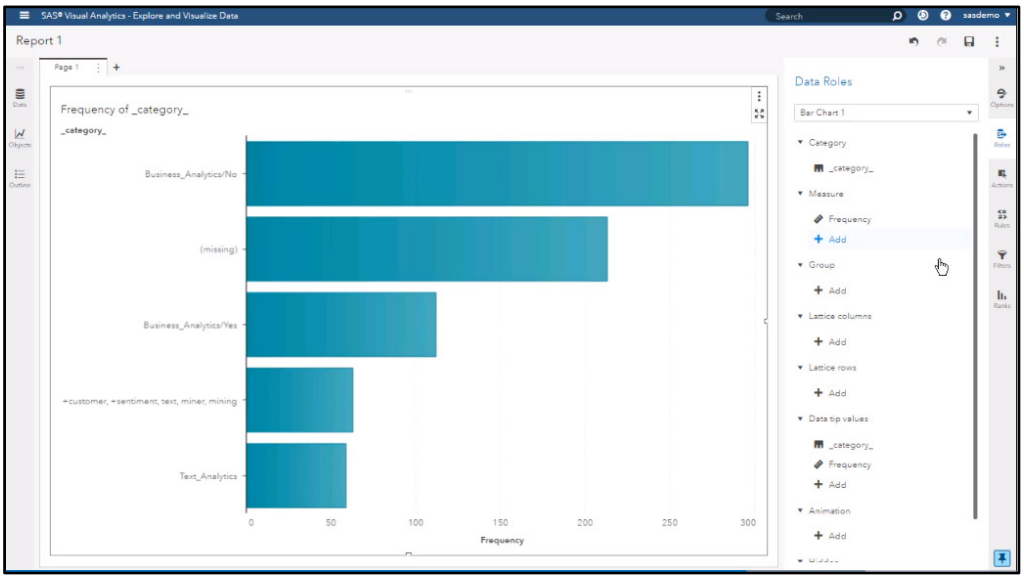


Bar Chart

Let's use a horizontal bar chart to visualize the categorical variables membership. Select Objects from the far-left pane. Under Graphs, select Bar Chart and drag it into the center window. On the far right of the graph window, under Category, click Add. In the Add Data Item pop-up menu, select__category__ as the category variable.

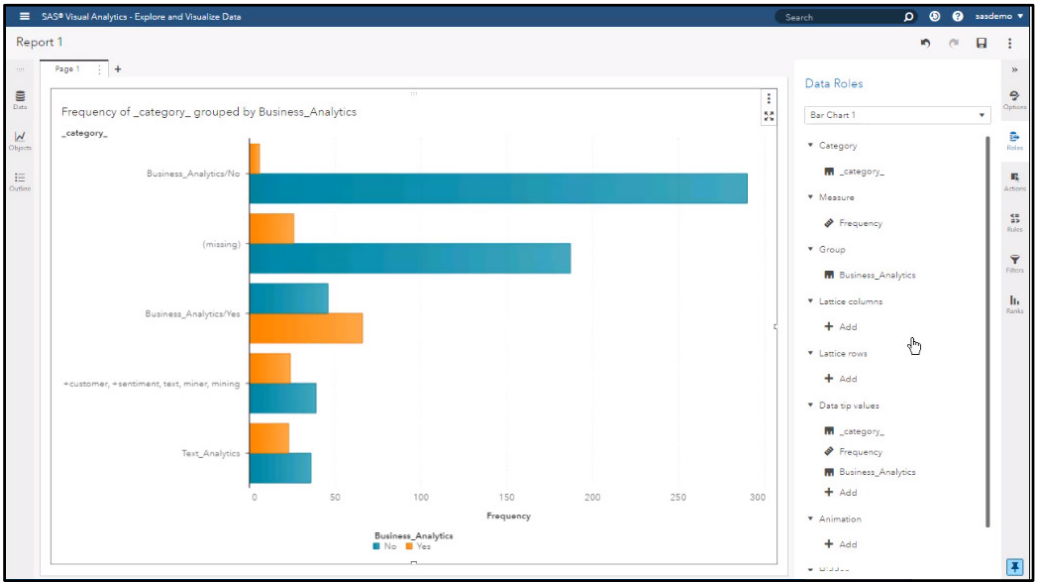
The bar chart displays a bar for each of the five subcategory levels, as shown in Figure 3.27. Notice that the subcategory levels are sorted by the categorical level frequency.

Figure 3.27: Frequency of _category_ Bar Chart



To generate a two-way bar chart, add a subgroup variable. Under Group, click Add. In the Add Data Item pop-up menu, select Business_Analytics as the Group variable. The chart now displays the frequency of _category_ grouped by Business_Analytics, as shown in Figure 3.28.

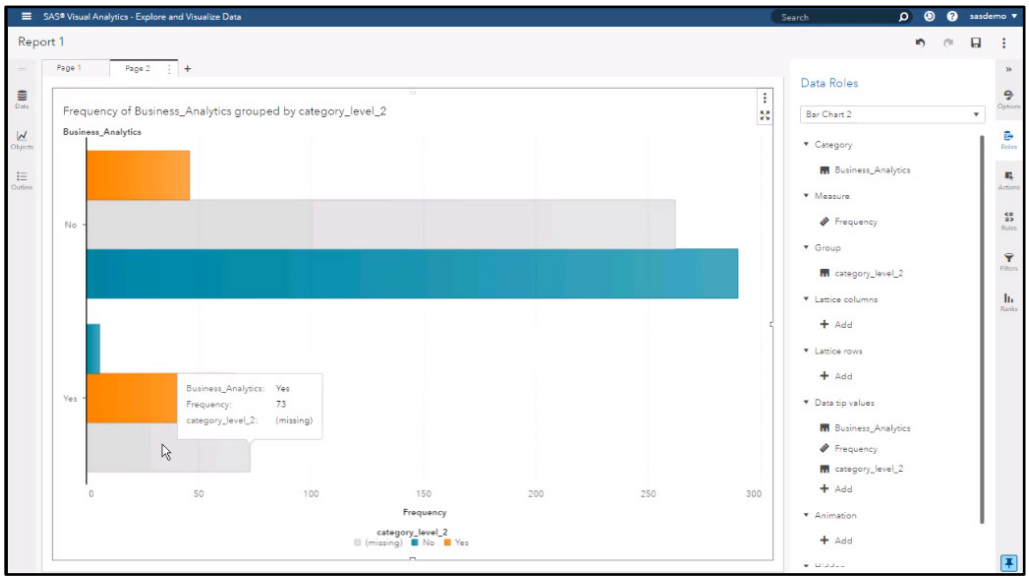
Figure 3.28: Frequency of _category_ Grouped by Business_Analytics Bar Chart



Similarly, we can create another two-way bar chart. Create a new page by clicking the + next to Page 1 on the tab at the top. Select Bar Chart from the Objects again, and drag it to the middle window.

This time, for Category, select Add and then choose Business_Analytics. For the Group, select Add, and then choose category_level_2. From this chart (Figure 3.29), we can estimate the model classification assessment statistics recall percentage (66 / (73 + 66 +6)) and precision percentage (66 / (66 + 46)) by hovering over the bars to see the numbers.

Figure 3.29: Frequency of Business_Analytics Grouped by category_level_2 Bar Chart

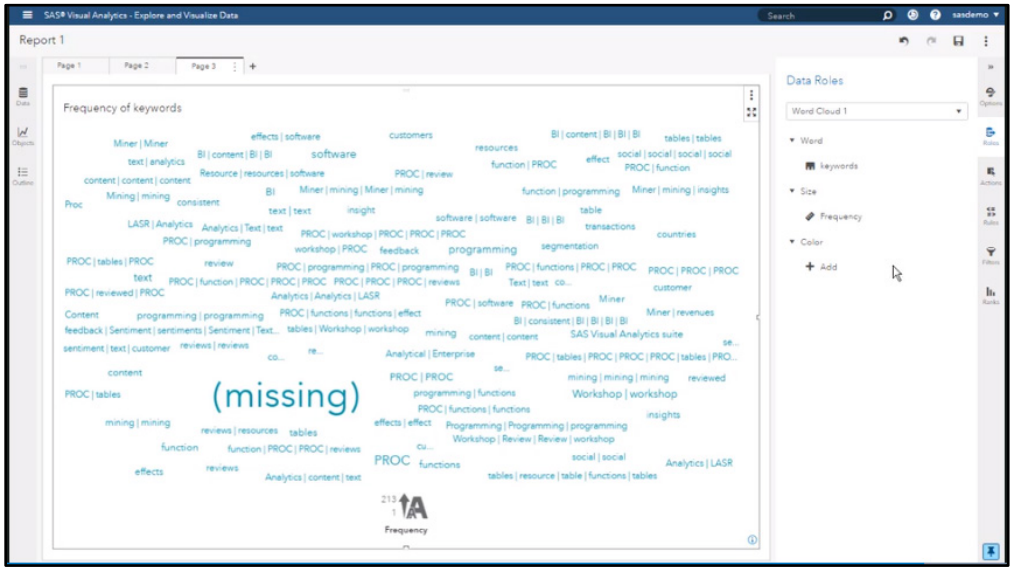


Word Cloud

To display the keywords frequencies used in the classification rules, add a new page. Then choose Word Cloud from the Objects list in the far-left pane. Drag the object to the center window.

In the right pane, for the Word variable, choose Keywords to be the variable. In the Word Cloud chart shown in Figure 3.30, the size of the font is proportional to the term frequency. In this example, the most frequent keyword is "missing," which represents the documents that do not belong to the predefined Business_Analytics category.

Figure 3.30: Keyword Frequency Word Cloud



To save your work as a SAS Visual Analytics report, click the Save icon in the upper right. Name the report "Category Analysis," and then click Save.

Because SAS Visual Text Analytics tools are integrated with SAS Visual Analytics, output tables generated by the SAS Visual Text Analytics nodes can also be explored in the following ways:

- Smart visualizations
- Guided analysis through auto charting
- Dashboard reports

Resources

This chapter is based on the "SAS Visual Text Analytics on SAS Viya" videos in SAS® Viya® Enablement, a free course available from SAS Education.

To learn more about SAS Visual Text Analytics, view screenshots and see other related materials, please visit sas.com/vta.

You might find the following documentation helpful as you learn more about programming in SAS Visual Text analytics:

- SAS Visual Text Analytics 8.3: User's Guide
- SAS Visual Text Analytics 8.3: Procedures
- SAS Visual Text Analytics 8.3: Programming Guide

Chapter 4: SAS Visual Investigator

Introduction 81
- Home Page 81
- Searches 82
- Workspaces 83
- Insights 84

Searches 85
- Simple Searches 85
- Advanced Search 86
- Geospatial Searches 92

Alert Triage Process 94
- Alerts Page 94
- Alerts Monitoring 96

Workflow Process 96
- Accessing Tasks 96
- Completing Workflow Tasks 97
- Task View 98
- Workflow Monitoring 98

Export to Microsoft Excel 98
Structured Printing 100
Audit Reports for Administrators 100
- Audit Search Query Builder 100
- Internal Entity Reference Number 101
- Data Object Import 102

Resources 102

Introduction

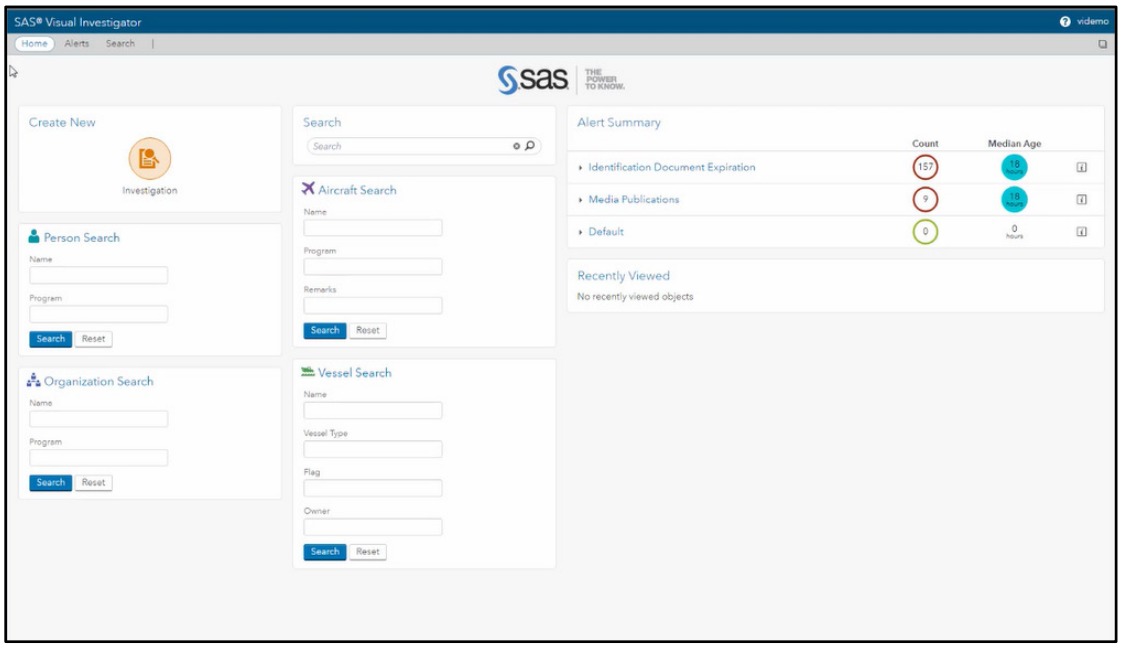
SAS Visual Investigator has a user-friendly interface that empowers intelligence analysts and investigators to work collaboratively and efficiently. Search capabilities enable you to build, gather, explore, visualize, and manipulate data. You can easily document findings by capturing views of search visualizations to narrate maps, timelines, networks, and other content. Auditing information related to user activities is captured automatically by the system and can be configured to meet your compliance needs. Let's take a quick look at some of the features, starting with the Home page. Then we will explore a few of the program's other features in more detail.

Home Page

From the SAS Visual Investigator Home page, you can access several capabilities. A sample Home page is shown in Figure 4.1. You can create new objects in the system, such as investigations, by using the Create New tile.

You can perform free text searches by entering search terms in the Search bar, or you can search for specific object types by using the form searches that have been configured on the Home page. These form searches provide you with quick access to objects that you need to work with frequently. For example, you could search for the name of a person or organization.

Figure 4.1: Visual Investigator Home Page



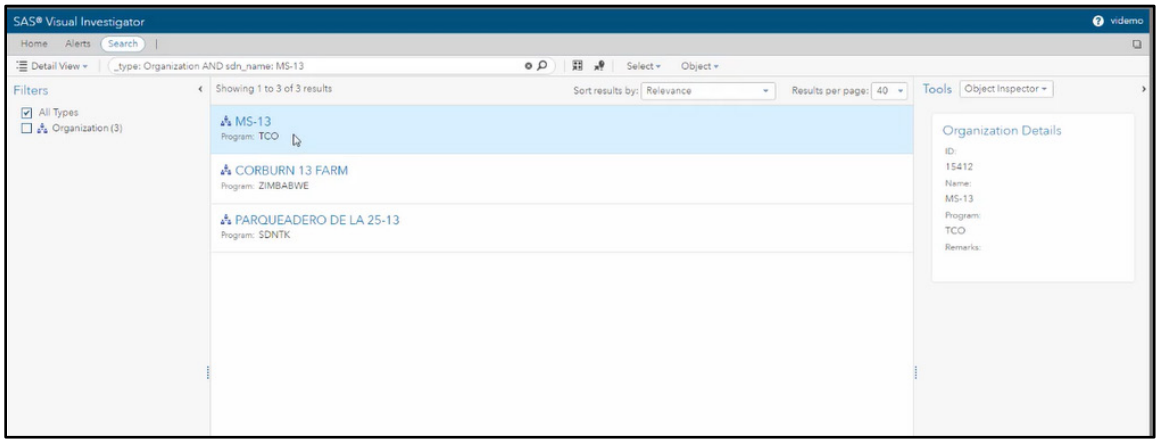
The Home page also provides you with an Alert Summary, which is an overview of your personal metrics, alerts, strategies and queues assigned to you, how many of them you have closed, and other actions. From the Recently Viewed tile, you can quickly access any objects that you have recently viewed.

Searches

Let's look at an example of how to do a form search. In this example, we are looking for information about a specific organization, so we will use the Organization Search tile on the Home page. Enter the search term "MS-13" in the Name field of the form search, and then press Search.

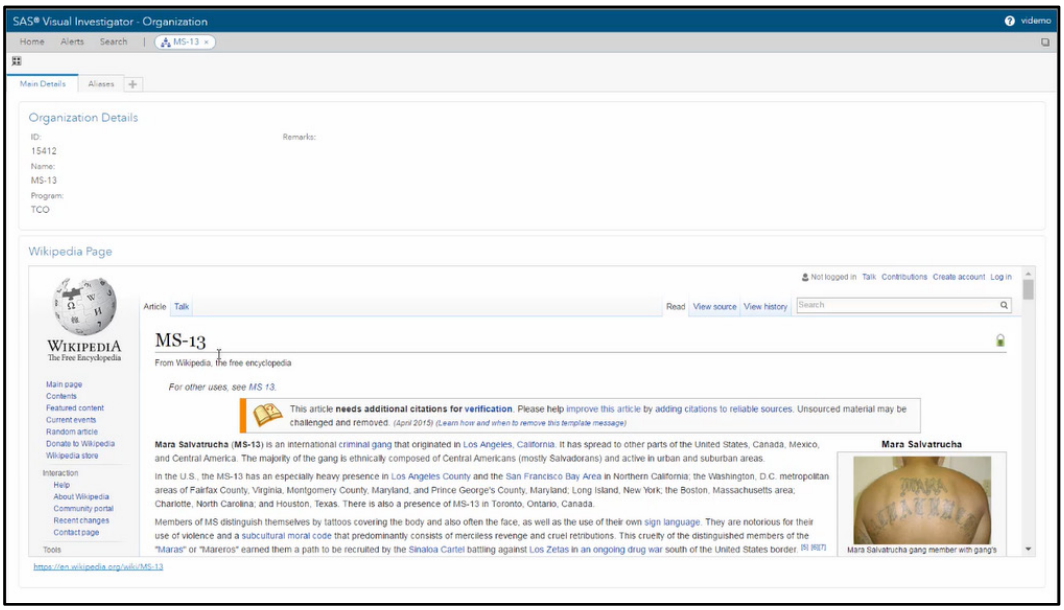
This will return search results, as shown in Figure 4.2. Use the Tools in the right pane to get more details about the results, including summary information.

Figure 4.2: Search Results



From the search results, double-click a result to open the object of interest. In this example, we will double-click the MS-13 result. A new window opens with the Organization Details for MS-13, as shown in Figure 4.3. Extra information is also available in additional tabs at the top, including details of the organization's aliases, and information from external sites such as its Wikipedia page might also be shown.

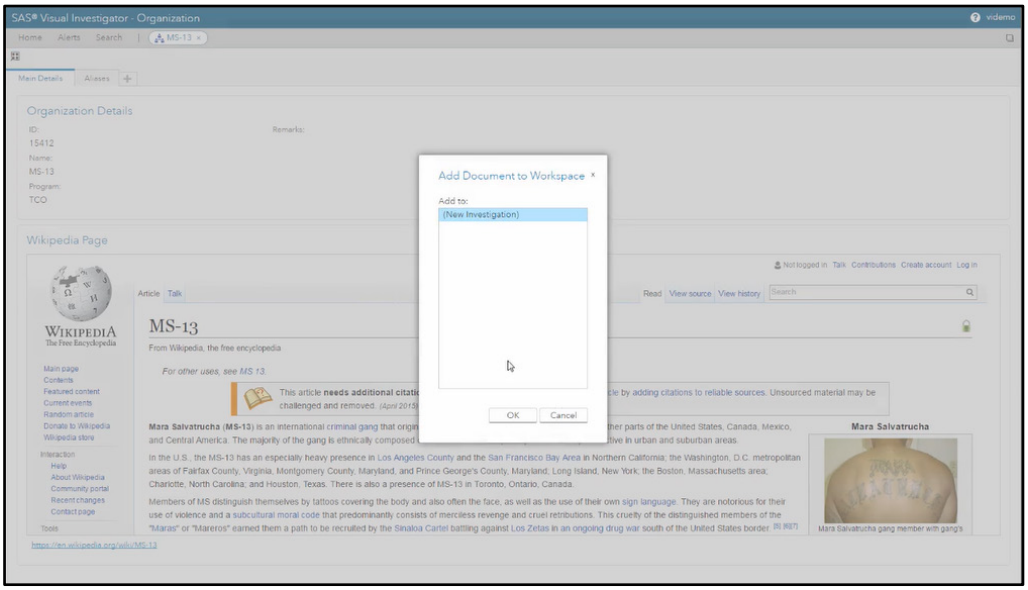
Figure 4.3: Organization Details Window



Workspaces

If you want to understand more about a search result's relationships to other individuals, places, and groups, you can add the object to a workspace and begin exploring the data in different visualizations. To add a search result to a workspace, click the square windowpane icon in the top-left corner of the screen. This opens the Add Documents to Workspace window, as shown in Figure 4.4. Select the option (New Investigation), and then click OK.

Figure 4.4: Add Document to Workspace

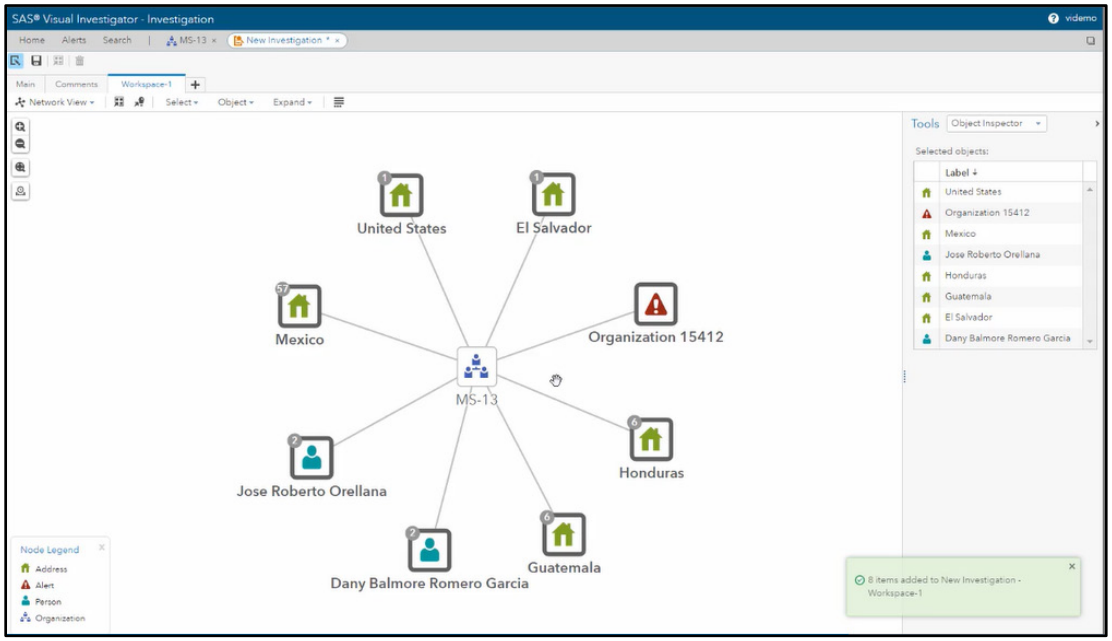


Now you will be taken to the workspace in your New Investigation and can explore different views. Click the Detail View drop-down on the left side of the screen to see your options for different views:

- Map View
- Timeline View
- Network View
- Table View

Let's see what the Network View looks like. You can expand the relationships to view the full social network, as shown in Figure 4.5. Double-click an icon to expand, or choose the expand option from the toolbar at the top of the workspace. You can flip out of this view easily by choosing one of the other views to see the Map View, Timeline View, or Table View at any time.

Figure 4.5: Network View



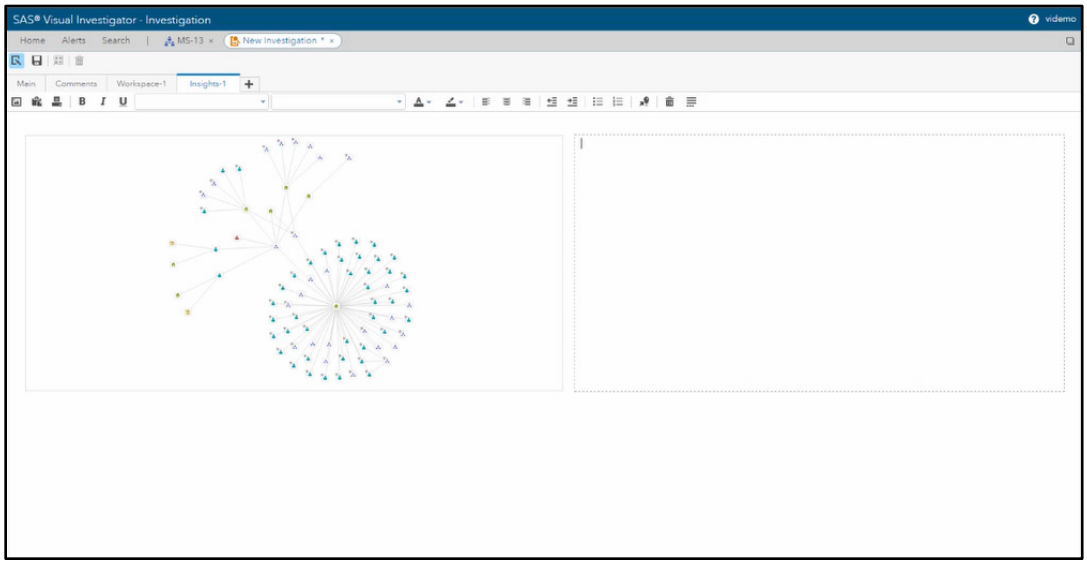
Insights

To capture any findings or items of interest during your research and investigations, you can add snapshot views of the visualizations and data to your insights and also add your own supporting notes.

To do this, select the Object drop-down from the toolbar at the top of the workspace, and then select "Add network to Insights." This opens the "Add network to Insights" dialog box. Select the Investigation and click OK.

Adding the network to your Insights opens the Insights tab, where you can continue building on your research, as shown in Figure 4.6.

Figure 4.6: Insights Tab



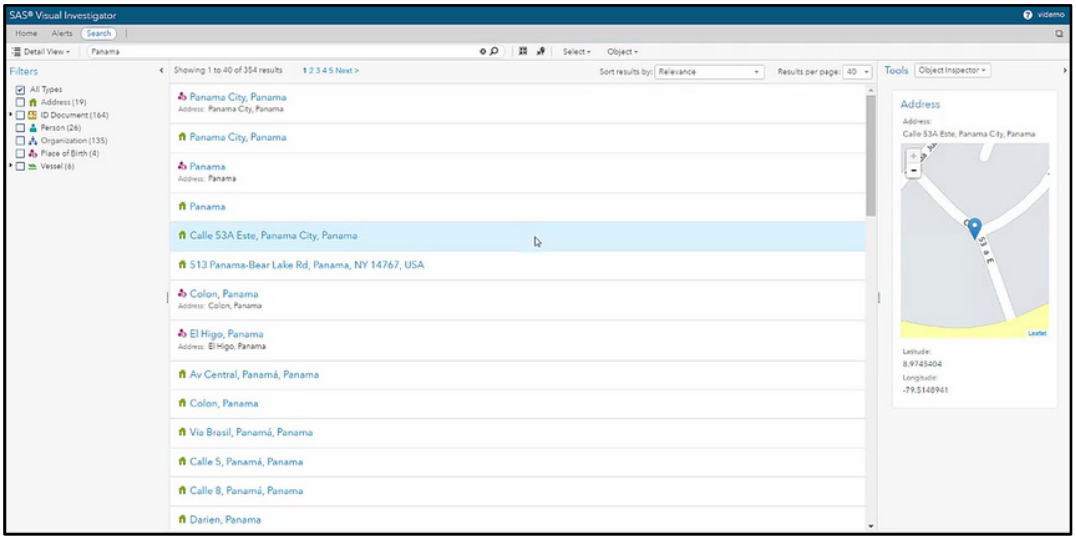
Searches

SAS Visual Investigator has the capability to perform simple free text searches from either the Search tab in the top toolbar, or the Search tile on the Home page. You also can perform advanced searches that enable you to create complex queries to further refine the results returned from a search. First, we will learn how to perform a simple search.

Simple Searches

Selecting the Search tab enables you to enter keywords and search terms. In this example, we want to look for information about Panama. Enter "Panama" in the search box and press Enter. The search results are displayed in a detailed tiled view, along with summary information and details of the number of results, as shown in Figure 4.7. You can page through the list of results by using the page navigator at the top of the window, change the sort order, and change the number of results displayed per page.

Figure 4.7: Search Results

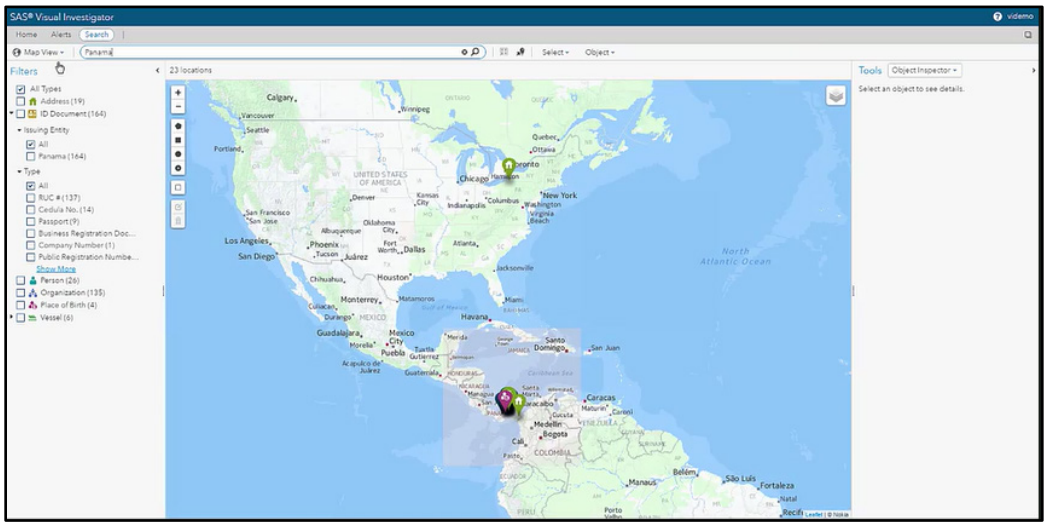


The Object Inspector pane on the right-hand side displays more information about a selected object. This enables you to see details and context without having to open each item individually.

The filters and facets on the left-hand side of the screen enable you to refine your search results. For example, you can click the check box next to ID Document to filter to show results for only a certain type of identification document, such as passports. These categorical facets, combined with date/time and numeric filters, can be used to further refine the results list.

You can also view your search results in alternative visualizations, including Map View and Table View. Click the Detail View drop-down list in the upper-left corner of the screen to change the view, as shown in Figure 4.8. Filters and facets are available in all views.

Figure 4.8: Map View



From any search result visualization, you can select one or more of the items in the results list and carry out an action on them by selecting them and right-clicking. For example, you can view the full details or add the objects to a Workspace or Insights as part of an investigation. All these actions are available from right-clicking on an object or from the toolbar menus.

Advanced Search

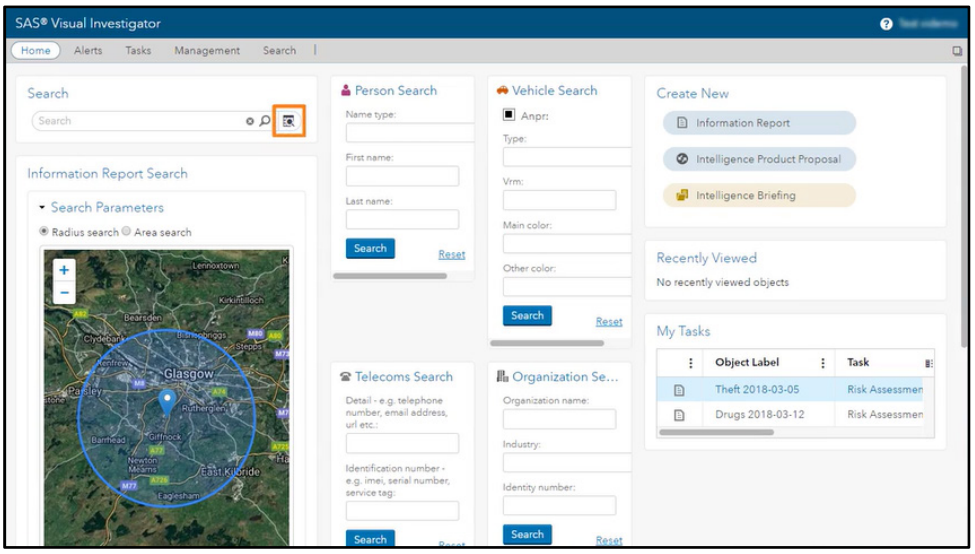
Now, let's learn how to configure and run an advanced search in SAS Visual Investigator. Advanced search provides a simplified means of creating complex queries to refine the search results returned from a search in SAS Visual Investigator. It does so by means of a simple-to-use query builder interface. Most users will not know the exact syntax to use to build complex search queries. Providing an advanced search with a query builder interface helps users easily define the query that they want to run.

In this section, we will look at an example of how to construct and run a search that will return people who have the first name of John and a last name starting with the letter S. We will then modify our initial search to add additional search terms.

Create a Query

To create an advanced search query, click the Advanced Search icon, which is located next to the Search field and is outlined in an orange box in Figure 4.9. Note that the advanced Search button is also displayed next to the search field on the Home page, the Search page, and in the search results.

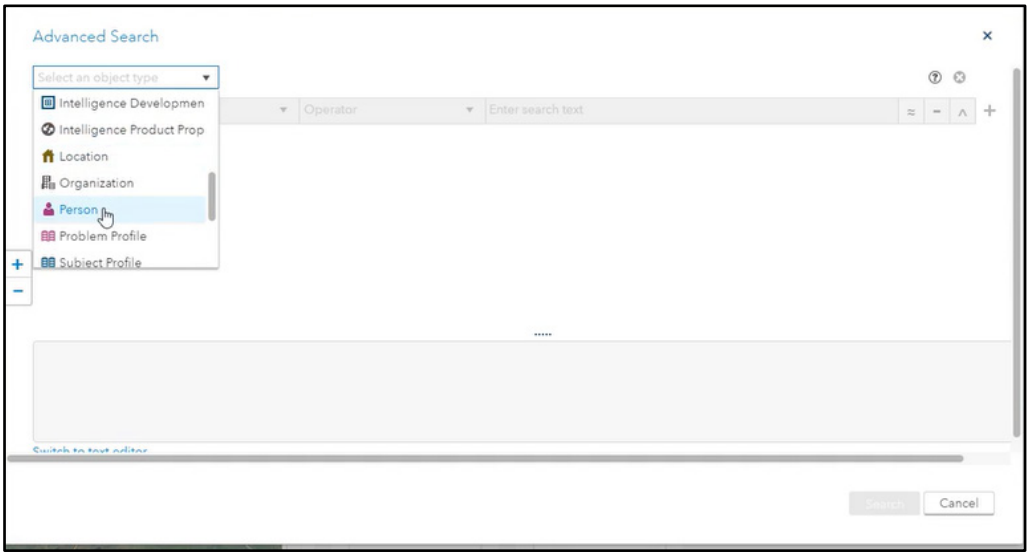
Figure 4.9: Advanced Search Button



After you click the Advanced Search button, the Advanced Search dialog box is displayed. The Advanced Search dialog box enables you to create advanced searches without knowledge of how to construct a query, using specific syntax for logical operators, fuzzy searching, proximity searching, and so on.

To begin creating your query, you must select the object type that you want to search for. From the Object Type drop-down list, select a specific entity type, such as an Alert, Information Report, Location, or Organization. In our example, we will select the Person object, as shown in Figure 4.10. Note that from the Object Type drop-down list, you can also select All Object Types if you do not want to choose a specific entity type.

Figure 4.10: Advanced Search Dialog Box



Selecting a specific entity enables you to create a query that targets specific information in a field, such as Last Name, when you select the Person entity type. Also, note that the object type selection that you make determines which fields and operators are available as you continue to build your query.

For this example, we will start by specifying the query parameters for Last Name. From the Category drop-down list, select the category that you want to run the query against. In this case, choose Last Name. From the Operator drop-down list, select the operator to be applied to your search. In this case, choose Contains all. In the Search text field, enter your search term. Our search term is S. As you begin typing, three query modifiers above the cursor that enable you to add wildcard and fuzzy modifiers, as shown outlined in an orange box in Figure 4.11.

Figure 4.11: Query Modifiers

We will apply a wildcard to our query. Select the wildcard modifier (*).

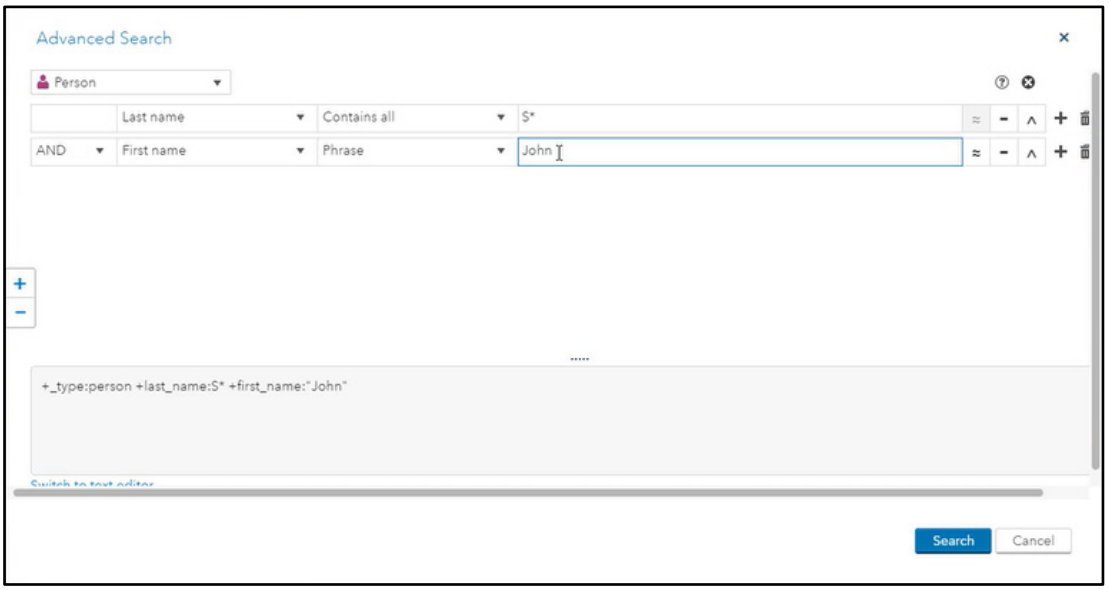
SAS Visual Investigator enables you to include wildcards in search queries to find similar words. These include the following:

- * – matches any number of wildcard characters (including no character) within a word. For example, *d*g* finds *dog, dug, dig,* plus *drug, drowning*.
- ? – matches a single wildcard character. For example, *d?g* finds *dog*, *dug*, or *dig*.
- ~ – enables fuzzy matching and can find objects that contain terms close to the search terms without matching exactly—for example, when words stored in the data repository are commonly misspelled or mistyped.

Note that as you build up your query, the text area in the gray box at the bottom of the screen displays the generated query text. You can apply modifiers to your search by clicking the appropriate button next to the Search text field. The Switch to text editor link at the bottom of the screen enables advanced users to enter queries manually without using the query builder tools.

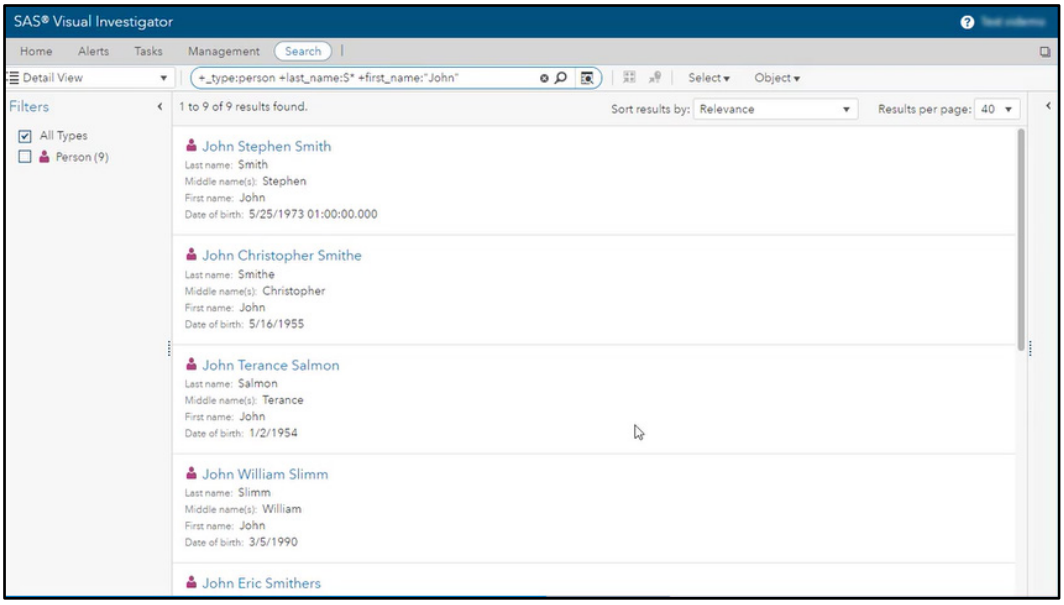
Our first query is complete, so let's add additional query parameters for First Name. Click the Add button (+) to the right of the first query to add a new row. From the Category drop-down list, select the category that you want. In this case, we will select First Name. From the Operator drop-down list, select the operator to be applied to your search. In this case, we will select Phrase. In the Search text field, enter your search term. We will enter the first name John, as shown in Figure 4.12. Note that you can click the trashcan icon to the right of the row at any time to remove the corresponding row from the query.

Figure 4.12: Additional Query Parameters



To execute the query, click the Search button. The query executes, and the search results list people who have the first name of John and any last name starting with the letter S, as shown in Figure 4.13.

Figure 4.13: Search Results



This search returned several results. Note that the query is displayed in the Search field at the top and in the browser session. If you open the advanced search again, it will be populated with your previous query. Queries are not saved between sessions, however.

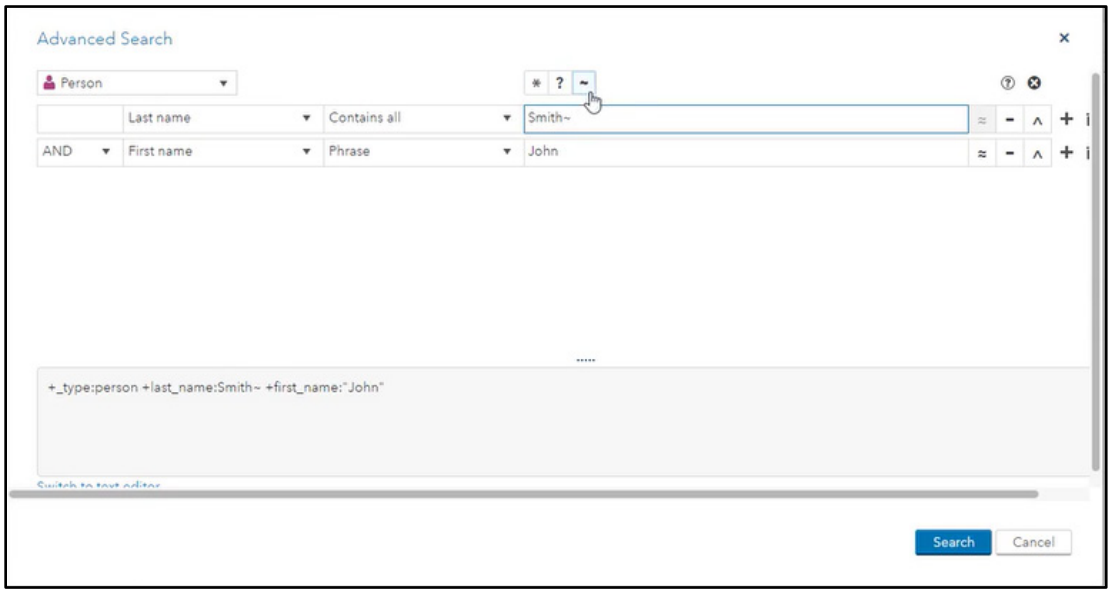
Modify a Query

What if you want to modify this query to add additional search parameters? Click the Advanced Search icon to the right of the search bar at the top of the screen. The Advanced Search dialog box is displayed, populated with our previous query.

Suppose we now want to search for individuals with the first name John and the last name Smith. However, we want the search to include alternate spellings for Smith.

To do so, first we need to modify our existing search criteria. Update the first row by changing S* to Smith. Add the fuzzy search modifier (~) after the word Smith so that we can run a fuzzy search, as shown in Figure 4.14. In this example, adding the fuzzy search modifier returns records for any near matches to the name Smith including Smith, Smyth, Smth, and on the like.

Figure 4.14: Fuzzy Search



In addition, we will add a second search query. This second search will use a proximity search to find Information Reports that contain information relating to a suspect's "black leather jacket." Proximity searches enable you to specify the number of terms within which your query terms must occur. This attribute is always used with a proximity level from 0 to 10.

Click the Add button (+) to the right of the last row in the current query. This will add a new row for our second search. In the first column, change AND to OR. From the Object Type drop-down list, select a specific entity type. In this example, we will select the Information Report object. From the category drop-down list, select the category that you want. In this example, we will use the (all categories) option. From the Operator drop-down list, select the operator to be applied to the list. In this case, we will use Phrase. In the search text field, enter your search terms: black leather jacket.

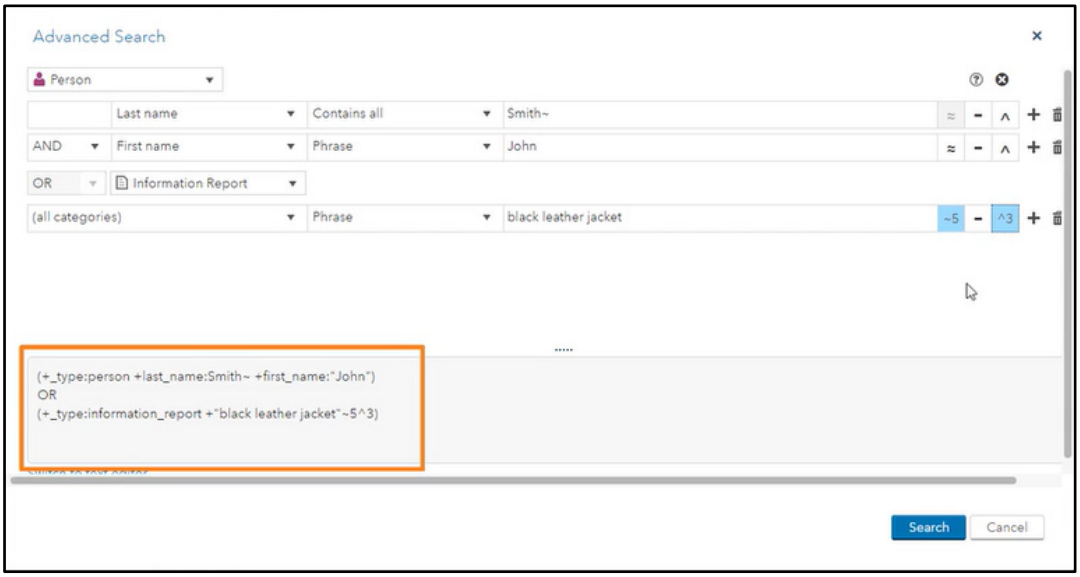
From the drop-down list accessed by clicking the ≈ symbol to the right of the row, select Proximity 5, as shown in Figure 4.15.

Figure 4.15: Proximity Search

This second search will return any Information Reports that contain the phrase "black leather jacket" to within five search terms. For example, this search would match with the phrase "leather jacket that was black."

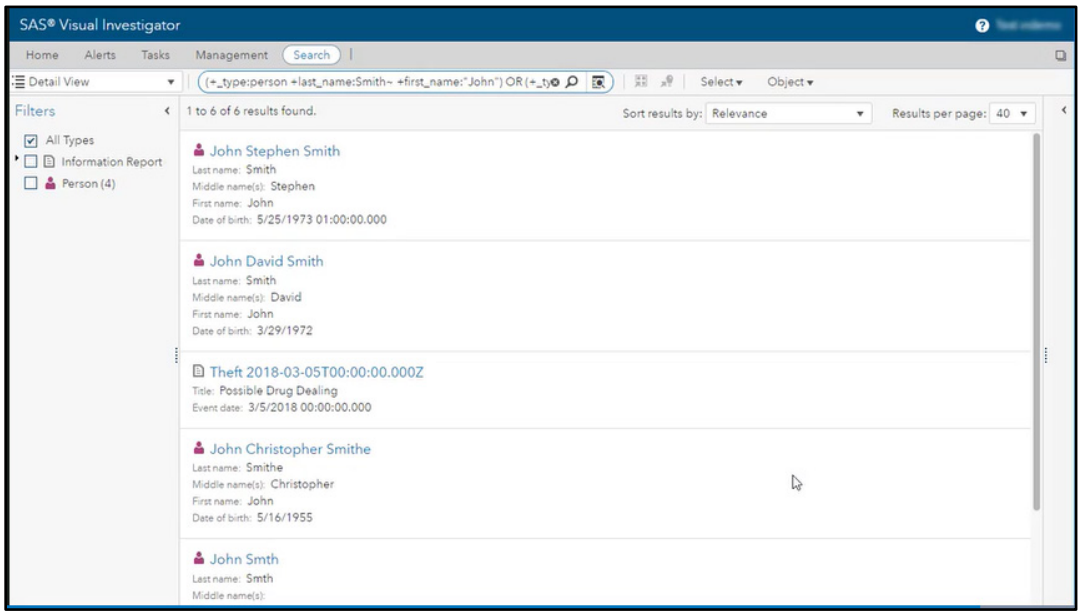
We will further extend our search to include a boost. The boost operator (^) is also located to the right of the row. It enables you to increase the importance of a word, phrase, or group within your search. We will increase the importance of the phrase "black leather jacket." Select the boost operator icon (^) and select Increase boost (3) from the drop-down list. The text area below updates to display your generated query text, as shown in Figure 4.16.

Figure 4.16: Generated Query Text



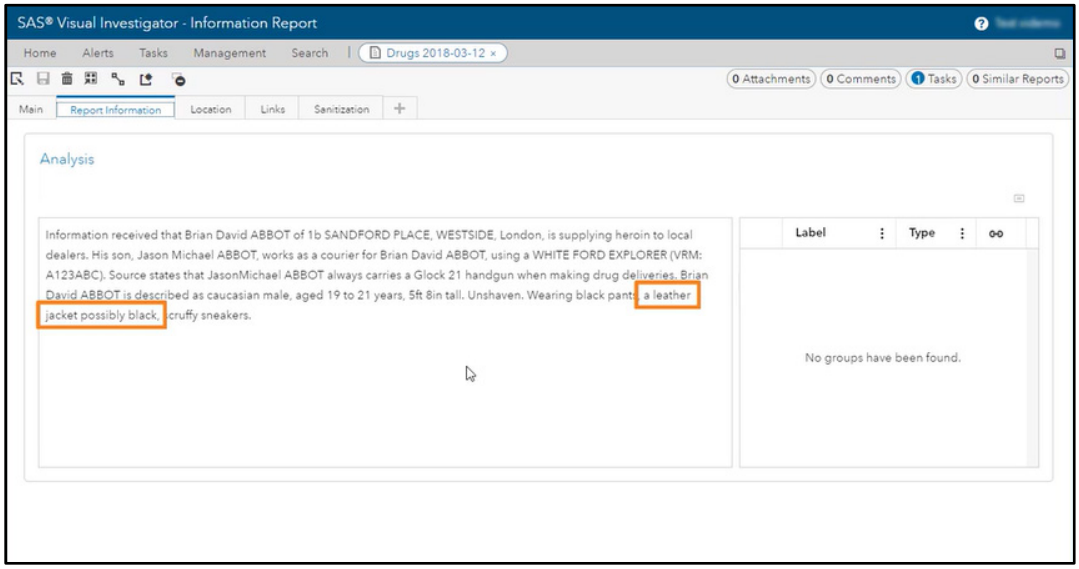
Click Search. Various search results are returned, as shown in Figure 4.17. Review the search results. Our modified search applies fuzzy searching to the last name field on the Person object, looks for the phrase John in the First Name field, and returns any hits.

Figure 4.17: Search Results



In addition, the second part of the search returns Information Reports that contain the phrase "black leather jacket" in the Structured Text field or in any records in which the terms black, leather, and jacket occur within five terms of each other, as shown in Figure 4.18.

Figure 4.18: Information Report Result



Geospatial Searches

SAS Visual Investigator provides you with rich geospatial search capabilities. From the Search window, you can switch to the Map View by clicking on the Detail View drop-down and selecting Map View. Then, you can zoom to the area that you want to search. You can zoom by using the mouse scroll wheel and the plus and minus bars on the Map toolbar.

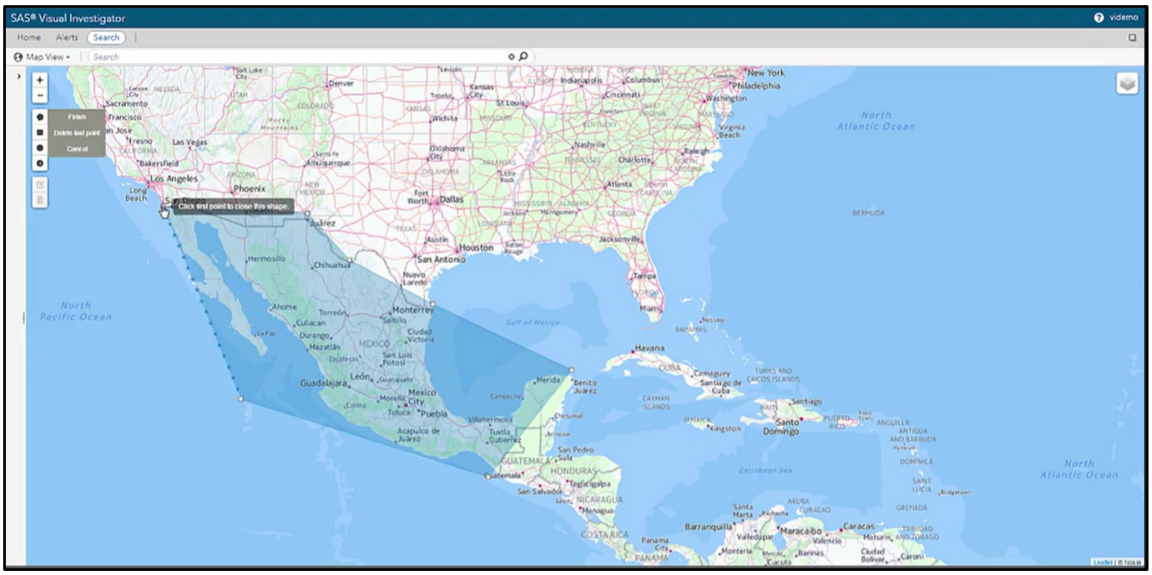
SAS Visual Investigator also provides some different area shapes that you can draw:

- Polygon
- Rectangle

- Circle
- Predefined radius

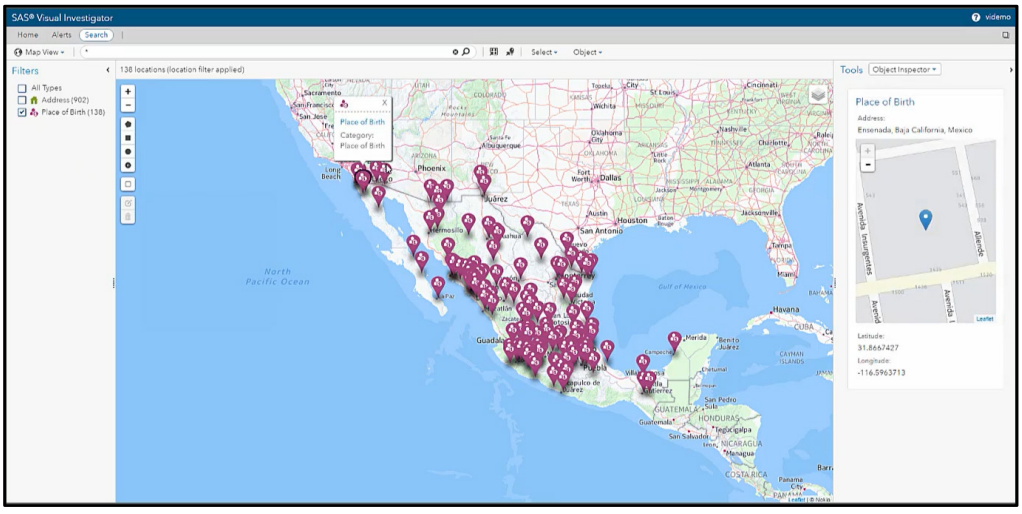
Suppose that we want to find more information about the data that we have in relation to Mexico. We can draw a bounded polygon shape roughly around Mexico, as shown in Figure 4.19.

Figure 4.19: Polygon Search Area



Once you have drawn your shape, then enter your search term in the search bar at the top of the page. In this example, we want to see all data in that area, so we will use the wildcard character (*). The results are plotted on the map. You can use the filters panel on the left to narrow down the results set, as shown in Figure 4.20. The Object Inspector pane on the right shows you more information about the results without your having to open each item.

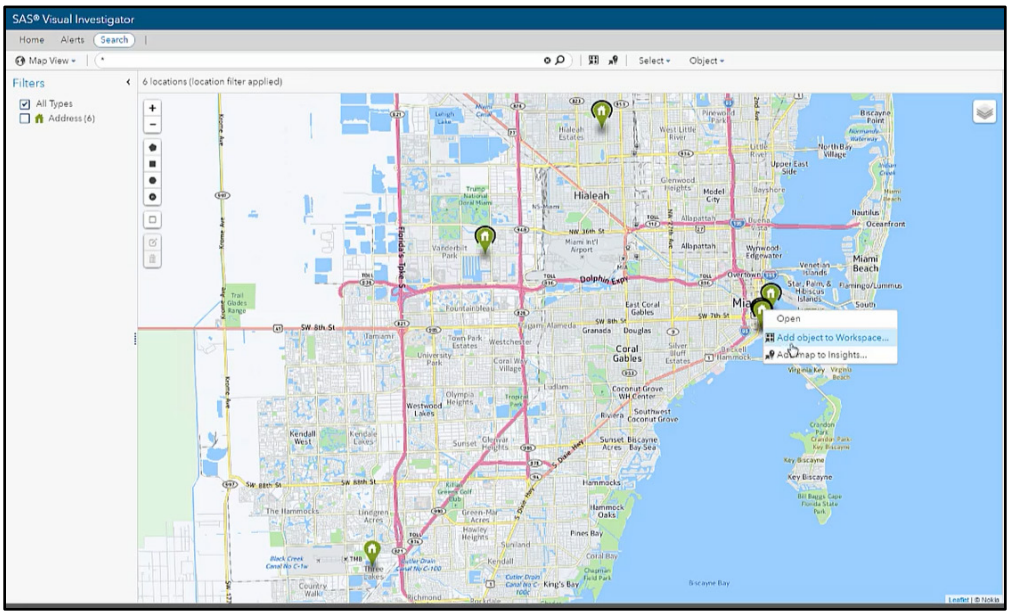
Figure 4.20: Search Results



Let's look at another example of how to perform a geospatial search. Move over to the Miami area on the map, and we will perform a search there. This time, we will use the predefined radius search option. For this example, we will choose a radius of 20 km from the drop-down menu and then drag the circle over to the map on which you want to perform your search.

You can use the selection option from the toolbar (open rectangle) to select multiple results within your search area. Then you can add them to a workspace for further investigation. Or you can add the Map View itself to Add Map to Insights, as shown in Figure 4.21.

Figure 4.21: Search Result Options

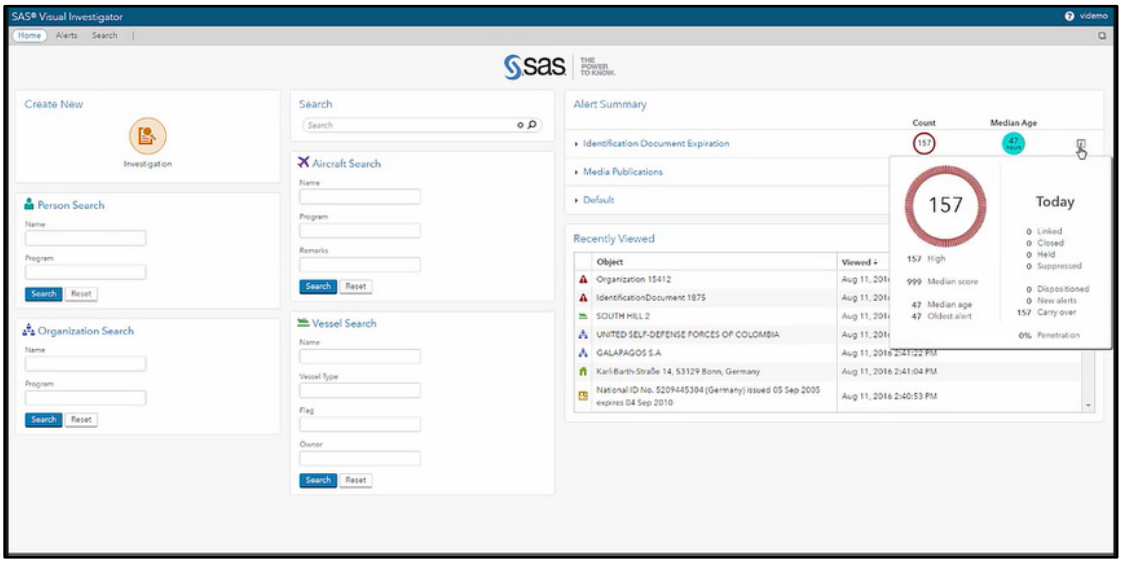


SAS Visual Investigator also provides the ability to search multiple areas simultaneously or edit a search area that you have drawn. Similarly, if you draw an area but change your mind, then you can delete it and draw a different shape.

Alert Triage Process

In this section, we will look at the alert triage process. From the Home page, you can see details of the alert strategies that you have been assigned to work, along with information about the number of alerts, their average age, and some personal metrics, as shown in Figure 4.22.

Figure 4.22: Alert Metrics

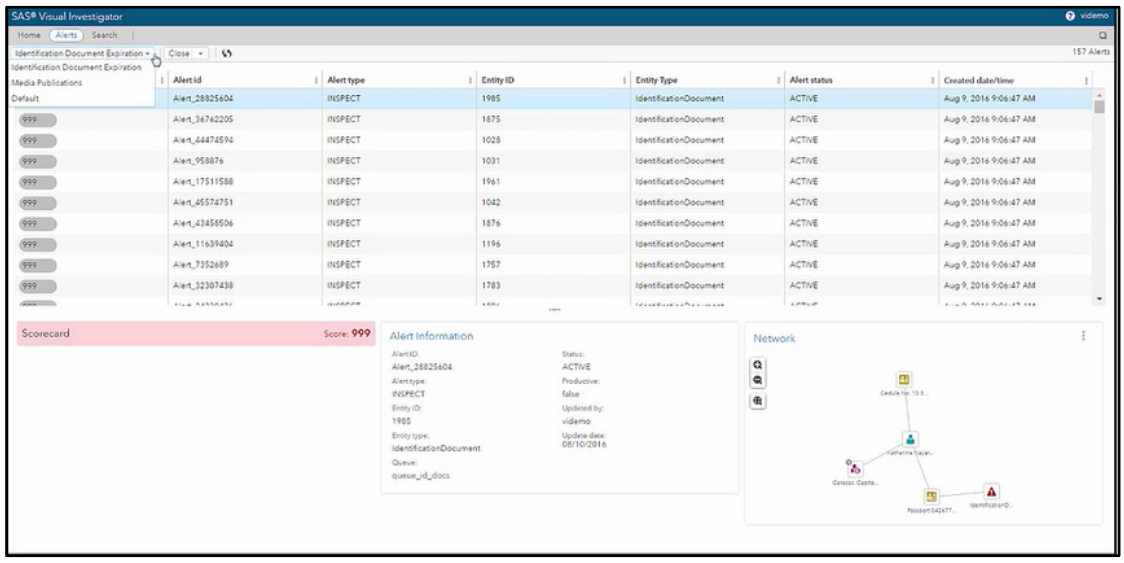


Alerts Page

From the main Alerts page, which can be accessed by clicking on the Alerts tab at the top of the screen, we can see details of all the alerts in a selected surveillance strategy. Alerts are triggered on the basis of rules and scenarios that identify events of interest that require triaging before a decision can be made on what action to

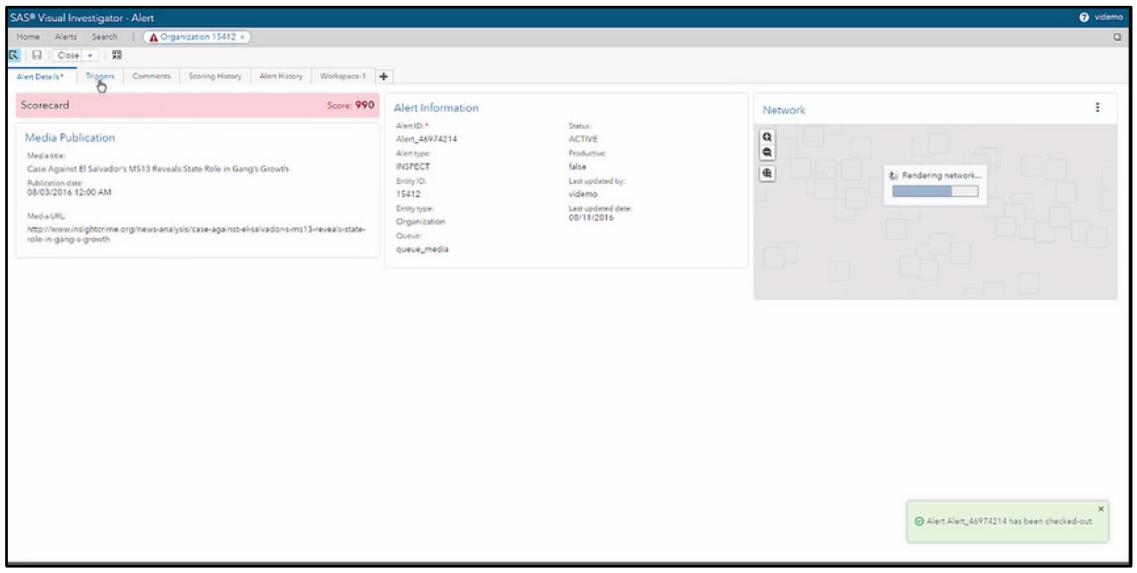
take. In Figure 4.23, we can see multiple surveillance strategies that have been deployed in the drop-down menu.

Figure 4.23: Alerts Main Page



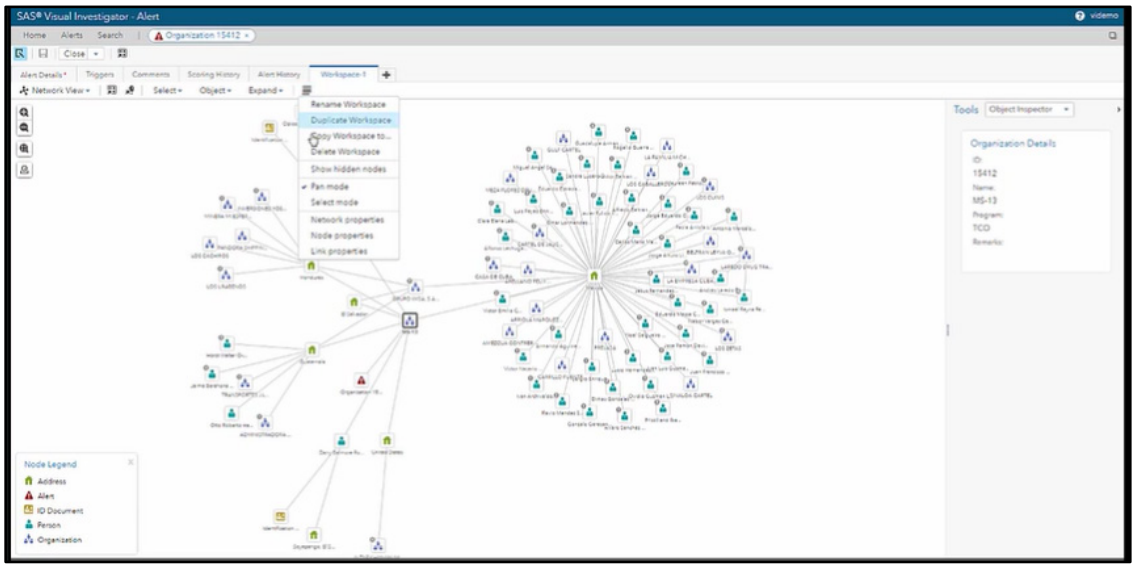
The first surveillance strategies show alerts raised in relation to individuals on watch lists whose identification documents, such as passports or diplomatic papers, are about to expire or have recently expired. The second surveillance strategy includes alerts that have been triggered because of individuals and organizations who appear on watch lists and are also referenced in recent news articles. From the Alert grid, we can see a summary page showing the scorecard information and details of the social network around the alerted entity. Double-click to open a specific alert and view its full details. In Figure 4.24, we see the scorecard and information about the rules that resulted in this individual's being sent to the Alert queue.

Figure 4.24: Alert Scorecard



We can also access the workspace that has been prepopulated with details of the alerted entity and its social network, as shown in Figure 4.25. You can explore the network by using the tools available in the pane in the right-hand side of the screen. You can also create duplicate or new workspaces and switch views to explore the information related to this alert and different visualizations, such as the Map View or Table View.

Figure 4.25: Alert Workspace



At any time, you can capture your findings in Add Map to Insights. Along with views of the data, you can also add images and your own notes to the Insights before deciding on the appropriate disposition action for the alert. For example, you might close the alert, put it on hold, or perhaps escalate it for further investigation.

Alerts Monitoring

Monitoring and reviewing activity status and details of user activity helps managers understand the activity associated with alerts, queues, and strategies. The Alerts Reports feature shows an overview of activities, as well as historical details. One of the benefits of this feature is its ability to provide information that helps organizations identify areas for improvement.

Workflow Process

The workflow functionality within SAS Visual Investigator supplements its alert triage capabilities by allowing resulting cases to be created and managed. Having both functions handled within a single application reduces the cost and support burden that comes with running multiple applications.

A user task within a workflow is something that is assigned to either a group or a user within SAS Visual Investigator. If the task is assigned to a group, then members of that group will be able to claim it. Claiming tasks prevents multiple members of the group from working on the same task at the same time.

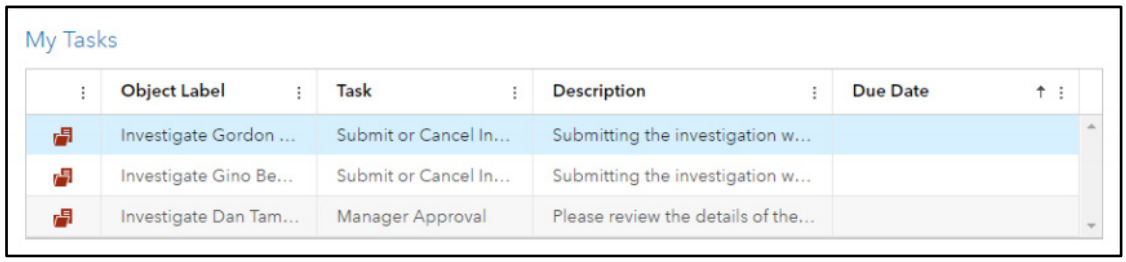
When adding a user task to a workflow, the designer can set the options that will be presented to the user performing the task. The user can then to mark the task as completed. For example, if the task is a review task, then the user might be presented with the following options:

- Accept
- Reject

Accessing Tasks

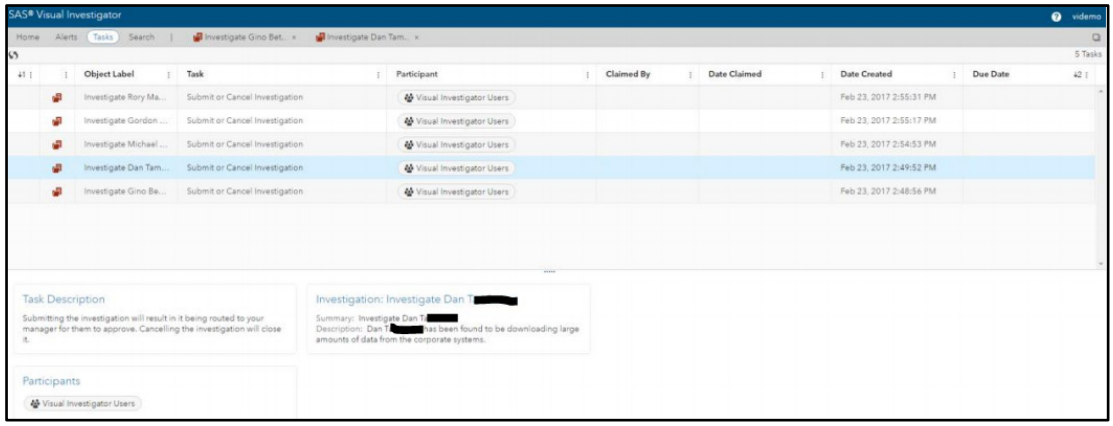
There are two ways that a user of SAS Visual Investigator can access his or her tasks. The first way is through the Home page. A Home page control enables users to quickly access the tasks that they have claimed (Figure 4.26).

Figure 4.26: My Tasks Section on Home Page



The second way that users can see the tasks is through the new Tasks tab. The Tasks tab enables a user to see all the tasks that are assigned directly to him or her, or to a group to which the user belongs (Figure 4.27).

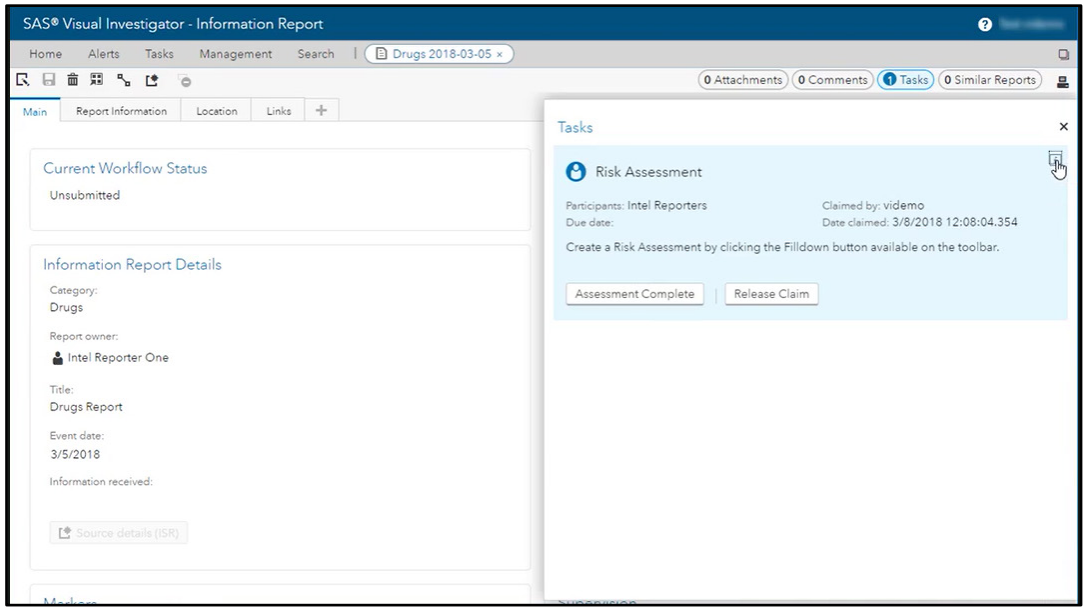
Figure 4.27: Task Listing within SAS Visual Investigator



Completing Workflow Tasks

Opening an object that has an associated workflow task will now result in a new toolbar button appearing. This button shows the count of tasks that are currently outstanding on the object. If the user clicks the button, then a pane will slide in to show details of the tasks, as shown in Figure 4.28. From within this task pane, the user will be able to claim the tasks. Once the tasks have been claimed, the options for completing the task will be available.

Figure 4.28: Tasks Detail

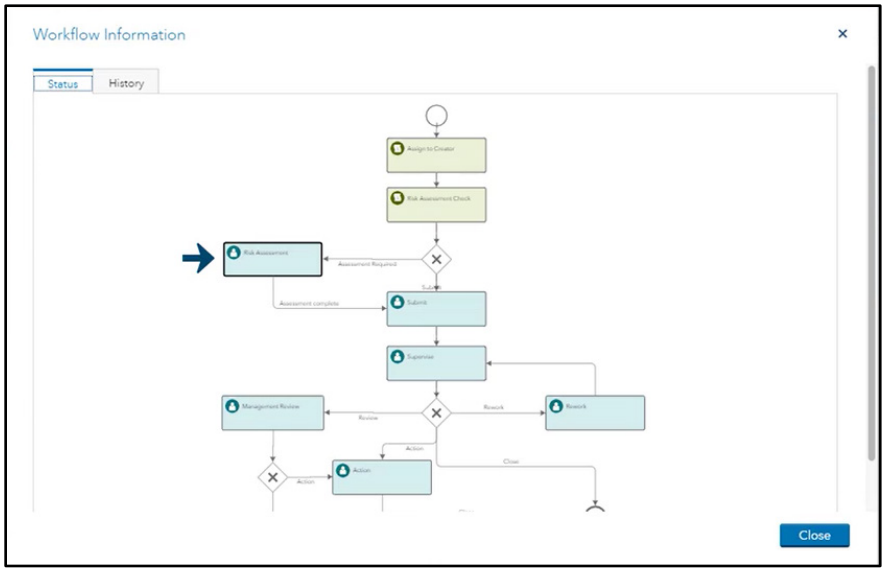


Task View

You can access the Task View in the workflow diagram by clicking the information button in the upper-right corner of the detail pane shown in Figure 4.28. This diagram makes it easier for users to understand where their task fits into the overall process and helps them understand what the next stage in the workflow is.

In this example, the Workflow Information Window in Figure 4.29 shows where the user's current task (highlighted with a bold border and arrow) fits in the overall workflow.

Figure 4.29: Workflow Information Window



The History tab displays the history of a workflow in which you are currently participating. You can view which users completed which actions and when these users completed their actions.

Workflow Monitoring

The Tasks Reports feature provides the ability for supervisors or managers to see an overview of completed and incomplete workflow tasks and review reports showing individual user activity. These reports show the times of peak activity for completion of workflow tasks and enable optimization of staffing.

Monitoring and reviewing workflow activity helps users follow the status of workflow tasks and usage activity in detail. This type of monitoring and review process helps identify areas for possible improvements and enables performance metrics to be reviewed in approximate real time.

Export to Microsoft Excel

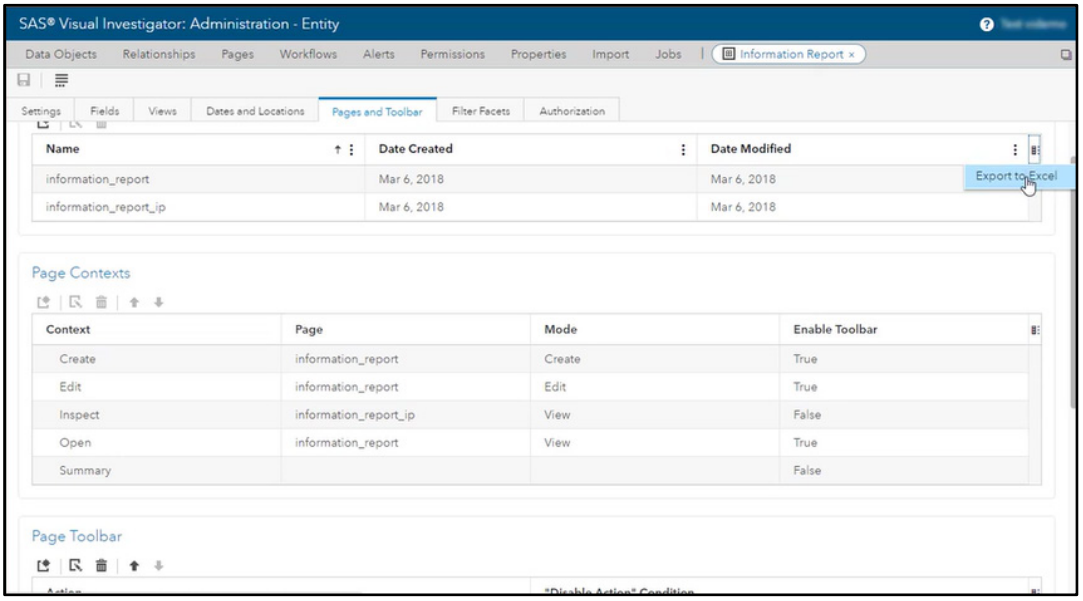
In this section, we will learn how to use SAS Visual Investigator to export data displayed within SAS Visual Investigator to Microsoft Excel files. Being able to export to Microsoft Excel files is a key requirement for analysts and investigators who need the ability to work with data in Microsoft Excel. This enables the data to be explored or shared outside of SAS Visual Investigator.

Exporting data to Microsoft Excel is standard on all grids, including the following:

- Alert triage grid
- Page builder grid controls
- Search results
- Table View
- Workspace Table View

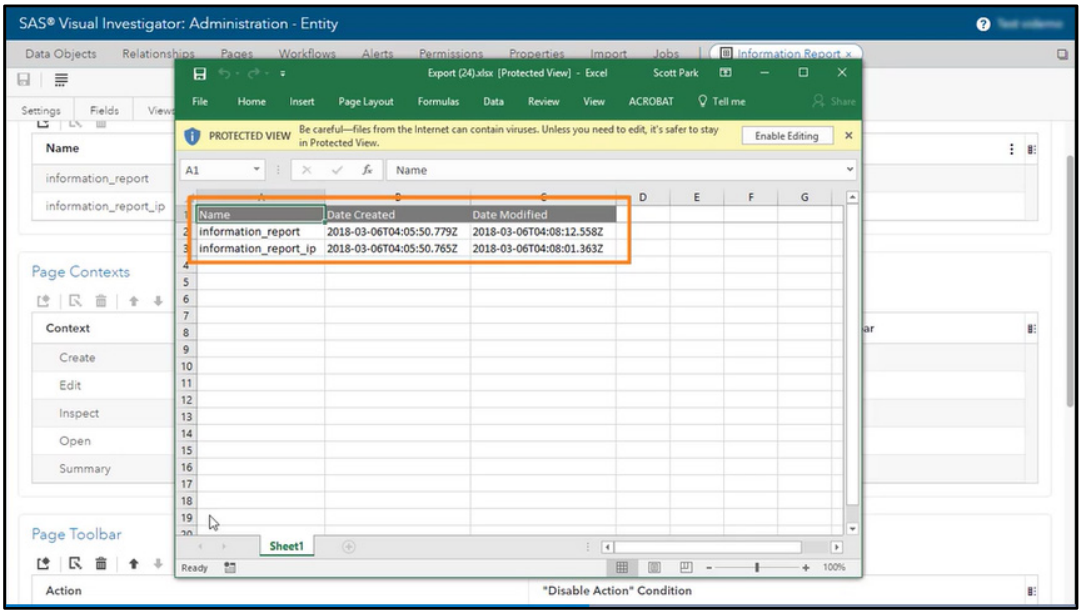
To export the data to Microsoft Excel, click the Export to Excel icon, and then select Export to Excel, as shown in Figure 4.30.

Figure 4.30: Export to Excel Option



A Microsoft Excel file is created with a separate sheet for the data from each grid on the page, as shown in Figure 4.31.

Figure 4.31: Microsoft Excel File



Administrators can limit the number of rows that can be exported to Microsoft Excel.

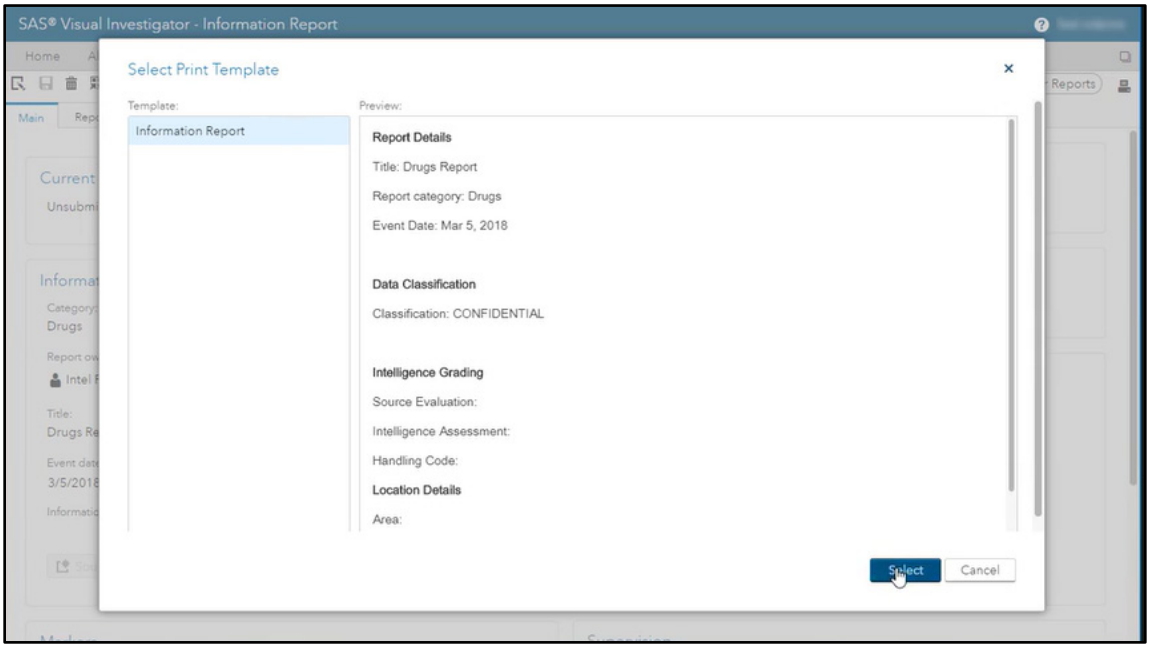
SAS Visual Investigator can also export the data from multiple grids on a page template when your application administrator has enabled the button on the toolbar. This export produces a Microsoft Excel file containing multiple tabs, one for each grid exported. For example, if a Person object contains multiple grids, with each grid detailing a particular aspect of an individual's features, such as facial features or marks or scars, then a Microsoft Excel file is created containing a worksheet for each grid.

Structured Printing

In this section, we will learn about user-access structured printing, a new feature that gives users the ability to print data in a structured form to comply with regulatory requirements. Paper forms often need to be re-created in the electronic world. For example, witness statements or the details of a subject profile investigation or case might need to be laid out in a certain way with the organization logo, headings, and fixed text.

Print templates enable users (with the appropriate permission) to print structured data from an object. Print templates are created and made available to users by the administrator on the basis of group memberships and access permissions. An example of a print template is shown in Figure 4.32.

Figure 4.32: Print Template Example



Users who are presented with the print option can select a template on which to base the output and then, generally, select from a variety of output options. Users can do the following:

- Select a template that indicates the format of the desired output
- Export the selection as a PDF file
- Invoke a single operation to export to a PDF file and attach the current object

Audit Reports for Administrators

It is important for organizations to capture information about the user activities being carried out in the system. This information helps organizations validate user activities. This information might be necessary as part of an internal investigation. This section contains new functionality and performance for administrators, as introduced in version 10.3.

Audit Search Query Builder

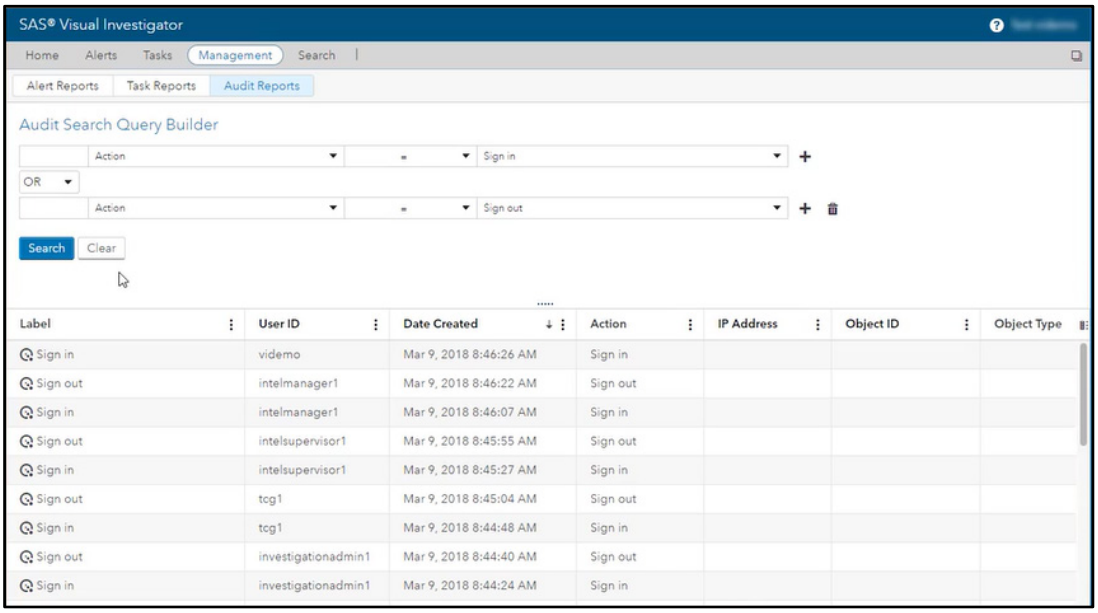
SAS Visual Investigator now captures more detailed auditing information related to key user activities within the system. Audited actions include (but are not limited to) the following:

- Signing in and out
- Searching (including details of the query and results, if desired)
- Viewing or editing an object
- Deleting an object

- Performing administrative actions, such as changing a configuration

Go to the Audit Search Query Builder on the Management tab, and then choose Audit Reports to start a new audit report, as shown in Figure 4.33. In this example, we are searching for any sign-in or sign-out action.

Figure 4.33: Audit Search Query Builder



The information captured in an audit typically includes the following:

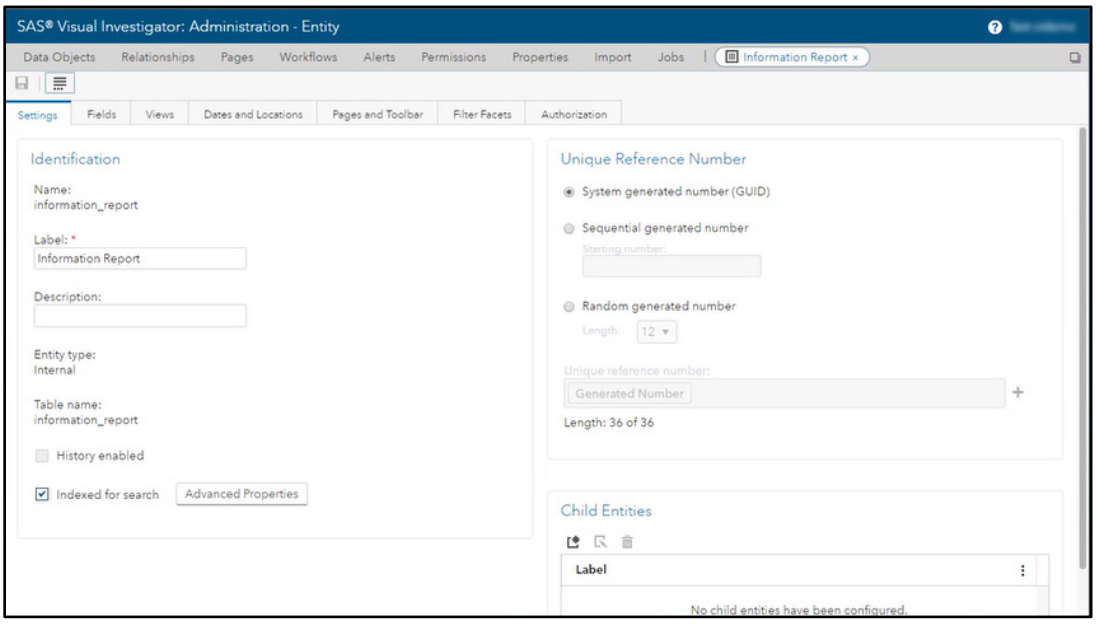
- User ID
- Date and time
- Action
- IP Address
- Object ID
- Object Type
- Additional information, as required, such as search query

Audit records are immutable and are created and managed by SAS Visual Investigator. The auditing level can be configured to ensure that the level of audit information captured is appropriate for a given customer.

Internal Entity Reference Number

SAS Visual Investigator enables internal entities to be configured to create a unique reference number that is human-readable and understandable by analysts and investigators. In the Unique Reference Number area in the Information Report settings, as shown in Figure 4.34, SAS Visual Investigator provides the ability for internal entities to be configured to enable a system-generated number (GUID, that is *globally unique identifier*) to be used. Alternatively, you could choose to create a Sequential-generated number or a Random-generated number to make the GUID more human-readable and understandable. This configurable format includes a numerical element and can include an administrator-defined string (that is, to represent the entity type) to include values from reference data in the entity.

Figure 4.34: Information Report Settings



Data Object Import

The Data Object Import feature provides a way for administrators to import a flat file or Microsoft Excel file into the product, or to connect to an external database to add entities (rather creating them manually). This feature now supports the ability to resume an incomplete import. An example of the use of this feature would be times that the import was interrupted by a session timeout or closing of the browser tab.

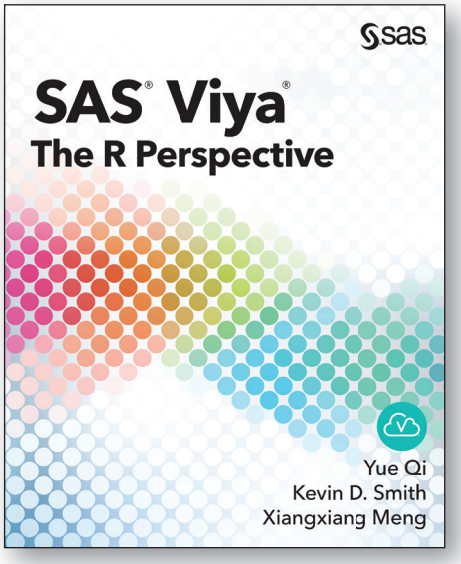
Resources

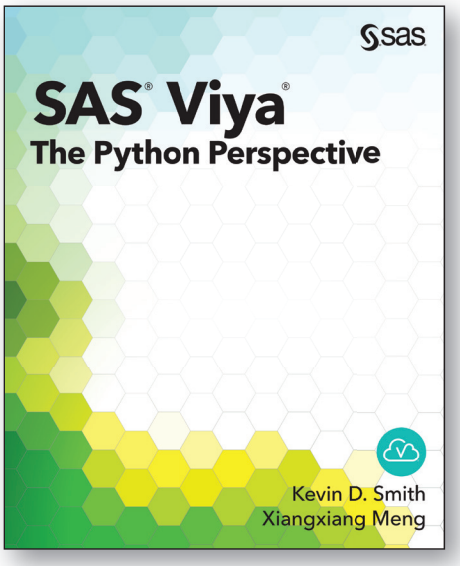
This chapter is based on the "SAS Visual Investigator on SAS Viya" videos in SAS® Viya® Enablement, a free course available from SAS Education. Some portions of the text were also based on the paper "Adding a Workflow to Your Analytics with SAS® Visual Investigator" by Gordon Robinson and Ryan Schmiedl.

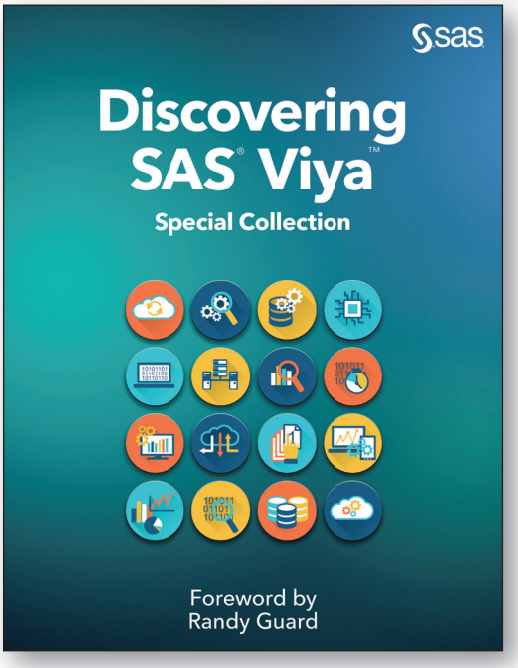
You might find the following documentation helpful as you learn more about programming in SAS Visual Analytics:

- SAS Visual Investigator Documentation
- Free tutorials and how-to videos

For more information on this topic, check out the books below in the SAS® bookstore:







For 20% off these e-books, visit **sas.com\books** and use code **WITHSAS20.**

sas.com/books
for additional books and resources.

sas
THE POWER TO KNOW.®

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. © 2019 SAS Institute Inc. All rights reserved. M1913158 US.0419