# Unlock the Power of SAS® Visual Analytics Starting with Multiple Microsoft Excel Files

Beena Mathew and Michelle Wilkie, SAS Institute Inc., Cary, NC

# **ABSTRACT**

SAS® Visual Analytics is a unique tool that provides both exploratory and predictive data analysis capabilities. As the visual part of the name suggests, the rendering of this analysis in the form of visuals (crosstabs, line charts, histograms, scatter plots, geo maps, treemaps, and so on) make this a very useful tool. Join me as I walk you down the path of exploring the capabilities of SAS Visual Analytics 6.3, starting with data stored in a desktop application as multiple Microsoft Excel files. Together, we import the data into SAS Visual Analytics, prepare the data using the data builder, load the data into SAS® LASR™ Analytic Server, explore data, and create reports.

## INTRODUCTION

The purpose of this paper is to demonstrate, starting with data in multiple spread sheets, the following capabilities of SAS Visual Analytics 6.3:

- 1. <u>Data preparation:</u> SAS Visual Data Builder, which is a component of SAS Visual Analytics, enables users to prepare data by summarizing, joining, and reducing the data. I will demonstrate how users can prepare data for exploration and mining by using SAS Visual Data Builder.
- 2. <u>Data Exploration:</u> SAS Visual Analytics Explorer, which is a component of SAS Visual Analytics, is a drag-and-drop data interface that can be used to explore data by using interactive visualizations such as charts, histograms, and tables. I will demonstrate interactive visualizations and, in addition, I will also demonstrate how users can apply data analysis such as forecasting, scenario analysis, and correlation by using SAS Visual Analytics Explorer.
- 3. <u>Reporting:</u> SAS Visual Analytics Designer, which is a component of SAS Visual Analytics, enables users to easily create reports or dashboards that can be saved and viewed on multiple formats and devices. I will demonstrate how to drag and drop tables, graphs, and gauges to create a well-designed report. Furthermore, I will demonstrate how to add text, images, and controls to reports.

# **GETTING STARTED**

## Brief Description of the Data Set Used in the Demonstration

The Insight toy data set is the sales-related information for a fictitious toy manufacturer. Its customers are "vendors" who resell their toys. The data spans 6 years (2008 to 2013) and includes the countries USA, Canada, and Mexico. Each year's data is stored separately as an Excel spreadsheet.

For each transaction, we have:

- a. Information about the items sold (product brand, line, make, style, SKU)
- b. The sale value ("order total")
- c. Various related costs (distribution, marketing, product)
- d. Information about the sales representative (rating, sales target, actual to date, and so on)
- e. Geographic information
- f. Information about the vendors (rating, satisfaction, distance to nearest facility)

## **Visual Analytics Hub (The Entry Point)**

As you first log on to SAS Visual Analytics you will see the Visual Analytics Hub as shown in Fig. 1. The Visual

Analytics Hub is an integrated environment that gives you the ability to access all the capability of SAS Visual Analytics. In addition, it gives you the ability to access the content that you and others have created. The following are the main features of the Visual Analytics Hub.



Figure 1

- The application **bar** enables you to return to the home page and to access other parts of SAS Visual Analytics and other SAS applications integrate with the home page. You can access your recently created or viewed reports, explorations, stored processes, data queries, or other objects in your recent history. Buttons are displayed for each open application
- 2. The menu bar enables you to access task options, such as creating reports, exploring data, managing your environment or favorites, setting your view, and getting help on using SAS Visual Analytics. You can search all SAS content from the menu bar, and you can log off from SAS Visual Analytics.
- The <u>Create Content</u> area provides icons to let you quickly explore data, create a new report, or prepare data, depending on your role and the associated capabilities. Other installed SAS applications might add actions to the Create Content area.
- 4. The <u>My Content area</u> lists any explorations, reports, queries, tables, or stored processes that you have opened or created recently. It also lists any content that you have marked as a favorite or as part of a collection. Click **Browse** to explore folders to find a report, exploration, stored process, or table.
- 5. The <u>other content</u> area displays lists of explorations, reports, stored processes, queries, or tables that your system administrator has added for you. If your system administrator has not created any lists, then you will not see the Other Content area.
- 6. The **Common Actions** section provides an alternate way for you to access features and other installed SAS applications.
- 7. Link section provides link to the introductory video as wells as links to other pages you have bookmarked.
- 8. The <u>SAS Resources</u> section provides links to resources on the SAS website and to social media.

# **IMPORTING DATA TO VA**

Visual Data Builder enables you to import Excel data(\*.xls, or \*.xlsx formats), any delimited text files, SAS data sets, Oracle database tables, and Twitter streams to the Visual Analytics environment and load it into the in-memory LASR analytic server. When you import data files from your machine, you are limited to files that are 4GB or less. This limitation is introduced by the web browsers. You can access the Visual Data Builder from the Visual Analytics Hub by double clicking the **Create Data Query** button or by selecting the prepare data icon from the menu bar.

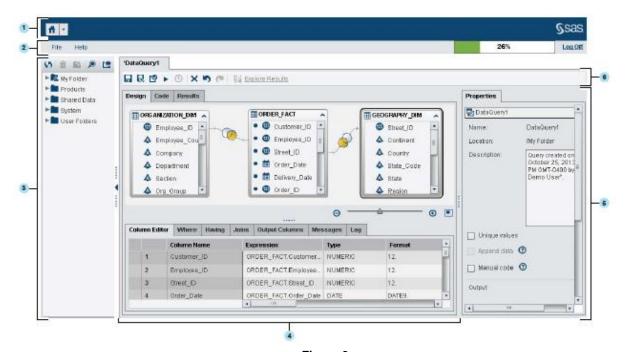


Figure 2

- The application bar enables you to return to the home page, access other parts of SAS Visual Analytics, and access other SAS applications that integrate with the home page
- The menu bar enables you to perform tasks such as creating new data queries and LASR star schemas. The
  right side of the menu bar has a memory gauge that displays the memory utilization for SAS LASR Analytic
  Server. You can also log off from SAS Visual Analytics from here.
- 3. The navigation pane displays a tree (the SAS Folders tree) of tables and data queries.
- 4. The center of the screen contains the workspace. When you create a new object, such as a data query, it is represented as a tab on the top of the workspace.
- 5. The bottom of the workspace contains a series of tabs that enable you to create column expressions and joins, and to filter data.
- 6. The properties panel enables you to manage the properties of the item that is selected in the workspace.

From File menu, choose **Import Data** and then **Local File** to select the file to import it directly on a SAS LASR Analytic server.

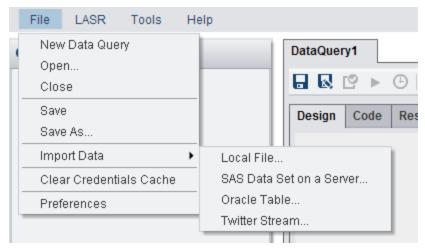


Figure 3

After selecting the files, the following window appears:

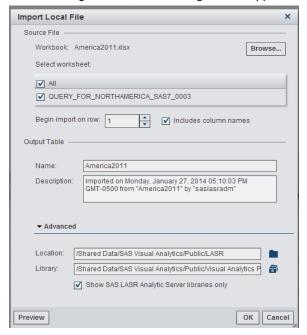


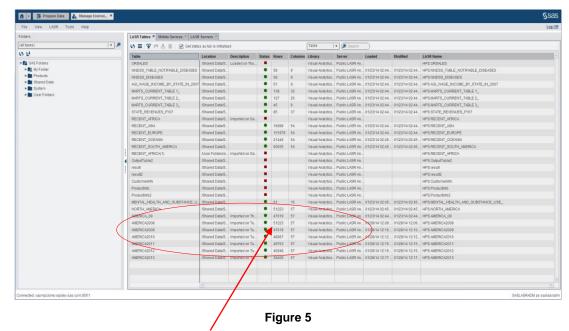
Figure 4

This window gives you the ability to perform the following tasks:

- Browse for the source file from your local machine
- Select the worksheet by checking the boxes
- Begin the import on any row using the down or up arrow to change the numbers
- Include the column name by selecting the check box when the row on which to begin the import has the column names.
- Change the name of the output table
- Select the LASR library and location where the data will be loaded
- Preview the column names and data values that will be available after the import using the **Preview** button

When you click **OK**, the data will be imported directly to memory on a SAS LASR Analytic server. You should be able to verify that the files are loaded to the LASR Analytic server by the following steps.

- Click the manage environment icon if from the menu bar or the **Manage Environment** link from under the common actions to open up the Manage environment window that is shown below.
- Select the LASR Tables tab
- Imported files will be shown with the status as a green circle that indicates the file is loaded. When you unload the data from the server, the status will be changed to a red square.



A green circle indicates that the tables are loaded to the LASR analytic server.

# APPENDING IN-MEMORY TABLES TO ANOTHER IN-MEMORY TABLE

We have imported several tables that represent different years of data . The next step is to append those tables together to get a historical perspective and plan for the future. The following are the steps to append the tables together:

• On the prepare data screen , click the LASR tab and select Append Tables



Select the base tables and source tables to the corresponding spaces as shown below. Here all the tables from year 2009 through 2013 will be appended to the base table America 2008.

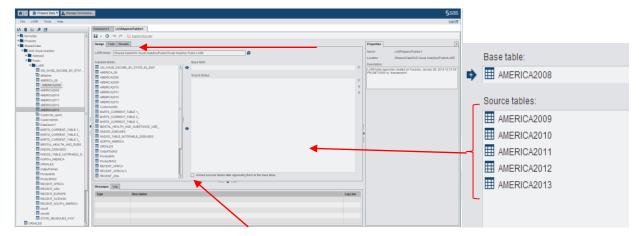


Figure 7

If you select the **Unload source tables after appending them to the base table** check box, the source tables are dropped from memory after the append completes. Only select this option if you do not want to access the source tables individually. Click the **Save** button, and specify a name and location of the query. Click **Run** to append the source tables. A pop-up window appears to show the results of the append query.

To preview the results, click **Yes** and the following window that displays the initial rows of the result table appears.

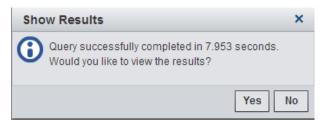


Figure 8

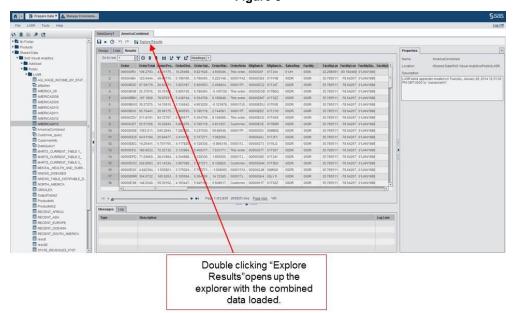


Figure 9

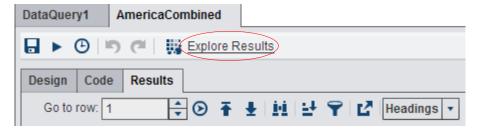


Figure 10

# **VISUAL EXPLORATIONS OF THE DATA**

The toolbar enables you to select

your visual explorations and visualizations.

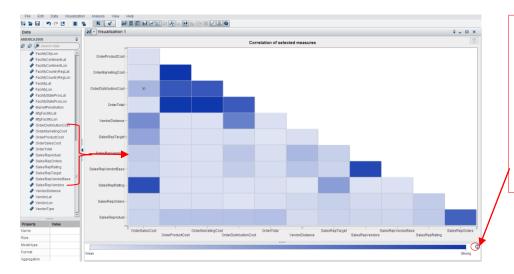
SAS Visual Analytics Explorer enables the user to perform interactive visualizations of the data using charts, tables, and histograms. In addition, the user can also apply data analysis such as forecasting, correlation, and fit lines.

The right pane Data panel contains lists the tabs that available enable you categories to change and the measures in properties your data of your visualizatio ns, filter the data, and create and view comments The workspace displays your visualizations

Figure 11:

# **WORKING WITH AUTOMATIC CHARTS**

One key technology that Visual Analytics Explorer has leveraged for ease of use is auto charting. This allows the user to see the data in ways they might not have considered, thus encouraging fresh new discoveries of data. The type of chart displayed is based on the data items assigned to it. For example, a measure data item will display a histogram, a category data item will display a bar chart, one geography and up to two measures will display geo map, three or more measures will display a scatter plot matrix or correlation matrix. The following correlation matrix is generated using the auto charting feature. The darker squares show that there is a strong linear relationship between the two variables.



About these correlation results displays the results of the charts in tabular form as well as a brief description of what correlation means. This allows the less statistically savvy user to be more innovative of the data

Figure 12:

You can see that the intersection of OrderTotal and OrderProductCost is a dark square with the correlation coefficient of 0.986, which indicates the strong linear relationship between those two variables. To explore further you can double click the square to explore the relationship in more detail. This detail is shown in the correlation regression chart below. The linear fit line is a straight line that best represents the relationship between the two variables.

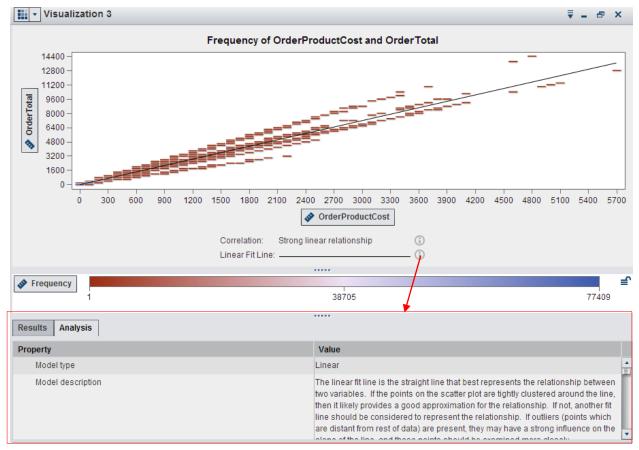


Figure 13:

## **CREATING GEOGRAPHY DATA ITEMS**

The following steps define a geography data item using a predefined geographic role:

- 1. In the Data pane or in the Data Properties window, locate a category that contains geographic information.
- 2. Change the role for the category to Geography. The Geography Role window appears.
- 3. Select the predefined geographic role for your category
- 4. Click **OK** to apply your changes.

If a predefined geography data item is not available, to create a custom geography data item the steps are as follows:

- 1. On the Data tab in the left pane, right-click the data item that you want to use for the geography data item. Select **Geography**, and then select **Custom**. The Geography window is displayed.
- 2. Select a measure for the **Latitude**. You can also enter the first letter of the name of the latitude column to quickly search for it in the drop-down menu.
- 3. Select a measure for the **Longitude**. You can also enter the first letter of the name of the longitude column to quickly search for it in the drop-down menu.
- Select a Coordinate Space. The choices are World Geodetic System (WGS84), Web Mercator, and British National Grid (OSGB36). The default is World Geodetic System (WGS84).
- 5. Click **OK**. The icon identifies the new geography data item on the Data tab.

The figure below shows how to convert the Facility City category into a custom geography item by making use of the Facility City Latitude and Facility City Longitude items available in the Insight Toy data set.

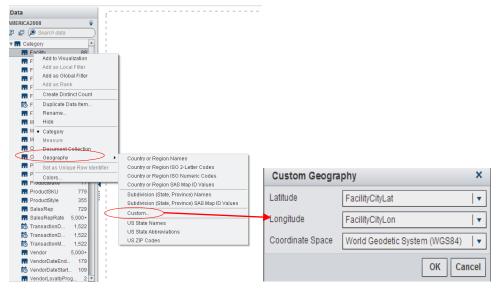


Figure 14:

#### **CREATING A NEW HIERARCHY**

Creating hierarchies enables you to add drill down functionalities to your data visualizations. The levels of a hierarchy are arranged with more general items at the top and more specific items at the bottom. There are two ways you can access these functionalities in Visual Analytics Explorer.

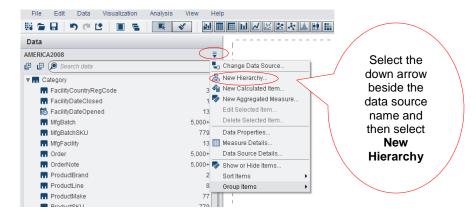
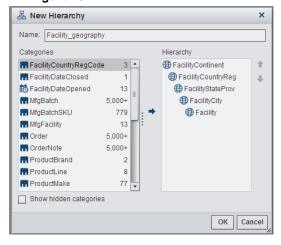


Figure 15:



In the popup window, enter the name of the new hierarchy and select the data items in the order of general to specific. Below is an example of the facility geography hierarchy from the facility continent to facility. Similarly, we can create another product hierarchy starting from the product brand ->product line->product make -> product style.

To change the order of the categories in your hierarchy, select the category that you want to move. Then, click 1 to move the category up, or click 1 to move the category down.

To remove a category from the hierarchy, select the category that you want to remove. Then, click —

Click **OK** to finish creating the hierarchy. The newly created hierarchy can be viewed among other data items as shown below.

Figure 16:



Figure 17

#### CREATING A NEW CALCULATED ITEM

On the Data tab, click , and then select **New Calculated Item**. The New Calculated Item window is displayed.

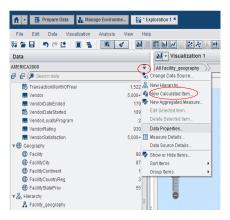


Figure 18

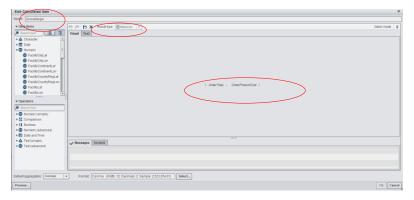


Figure 19:

Enter a name, select the result type, and build the expression for the item



Figure 20

Click **OK**. The new calculated data item appears in the Data tab. In the above example we have added one calculated item named Gross Margin = OrderTotal – OrderProductCost. Similarly another calculated aggregated item = (sum (GrossMargin))/ Sum (OrderTotal) is also created to use in our explorations.

# **ANALYZING GEOGRAPHIES WITH GEOMAP**

Select the Facility\_geography, and press the Ctrl key and select OrderTotal. Simply drag and drop them onto the work space as shown below. Because the category selected is a predefined item, Visual Analytics automatically displays a geo map visualization. Since we used a hierarchy, we can click the NA bubble to drill down to US and then to US states, and so on. We can also add another measure, GrossMarginRatio, to the workspace. The resulting visual is shown below. Here the bubble size represents the order total and the color represents the gross margin ratio. The hierarchy is drilled down to the US state level, which is indicated on the top of the screen. You can add your insights from this visualization as comments using the comment tab in the right panel.

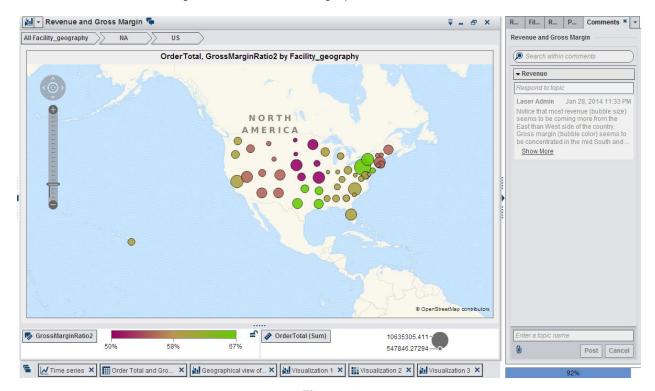


Figure 21

#### **CROSSTAB-GETTING STARTED**

A crosstab shows the intersection of category values and measure values. Starting with a blank workspace, select the crosstab from the toolbar.



Figure 22

This opens up a crosstab palette. Under the Roles tab on the right, you can add multiple categories and measures for Columns, Rows and measures.



Figure 23

From the Data tab, select Facility Country and drop it in the column box. Select Product Line and drop it in the row box and select OrderSalesCost and drop it in the measure box. You can also select the down arrows beside each box and select the data item from the list of items displayed. Now you will have the crosstab that shows aggregated Order Sales Cost for each Product Line by Facility Country/Region as shown below.

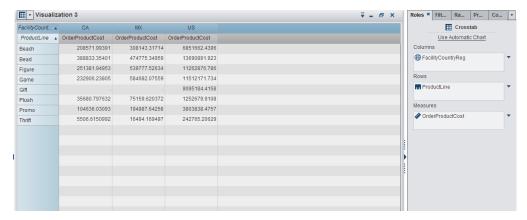


Figure 24

You can add depth to this table by adding the hierarchies to columns and rows. This time select Facility\_geography as the column and Product hierarchy as row. Because you are using hierarchies, you can drill down on any row and column as shown below.



Figure 25

Select the Properties tab on the right pane. You can turn on the column subtotals and totals as well as the row subtotals and totals. You can place the totals before or after the corresponding data item. You can also give a more meaningful name to the visualization.

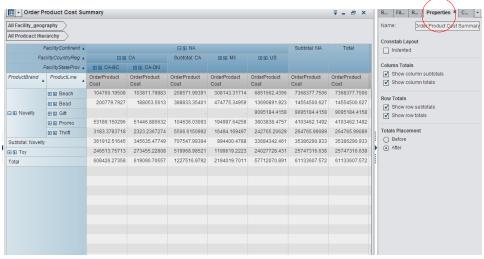


Figure 26

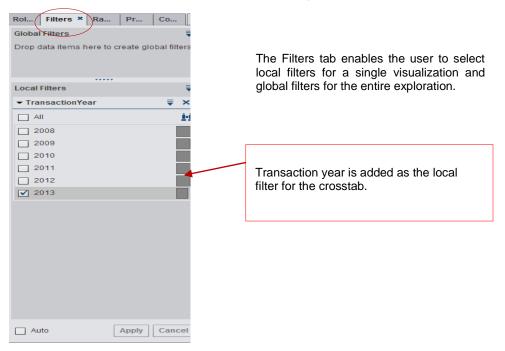


Figure 27

You can also add advanced filters as shown below to filter your data items.



Figure 28:

## FORCASTING -GETTING STARTED

Step 1 – Starting with a blank work space, select Line chart from the toolbar.



Figure 29:

Step 2 - From the Data tab select transactionMonthandYear and OrderTotal and drag them over to the workspace.

Step 3 - Turn on forecasting for this time series-based line chart to get the future predictions for OrderTotal.

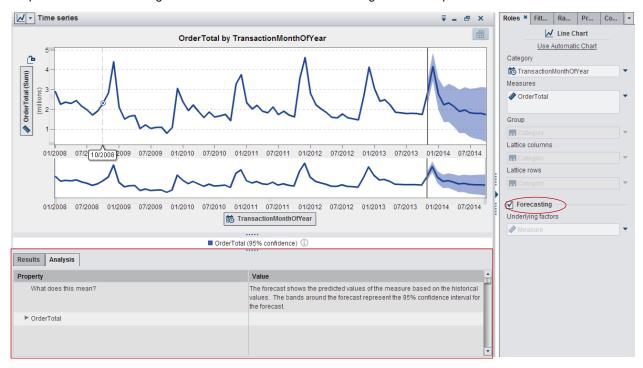


Figure 30

Visual Analytics takes a black box approach to analytics. There is not a manual way to modify the models. SAS will make decisions on what is appropriate given the data. This gives the business analyst who does not have a statistics background the capacity to takes advantage of analytics. Thereby, enabling the user to apply analytics in a business scenario and come up with interesting insights.

We can improve the accuracy of our forecast by adding the measures that we believe would have an influence on the forecasting. Select OrderMarketingCost, VendorDistance, and MarketPenetration and drag them to the **Underlying Factors** box.

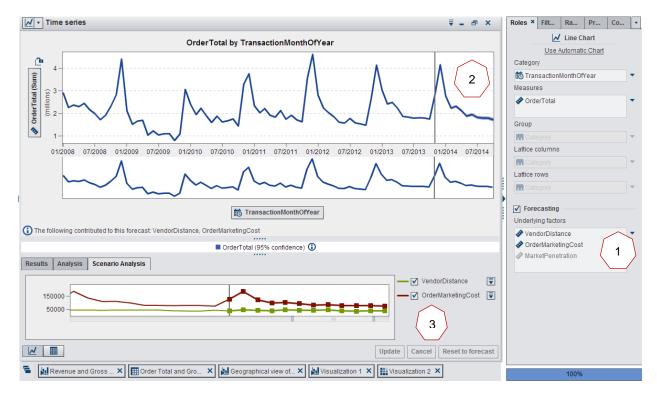


Figure 31

By adding those factors, the following changes have happened:

- Out of the three factors, MarketPenetration has been grayed out indicating it does not influence the forecast.
- The forecast itself has been recalculated and it is now more precise, which is indicated by the much tighter confidence interval.
- A new scenario analysis has been added allowing you to perform simulations.

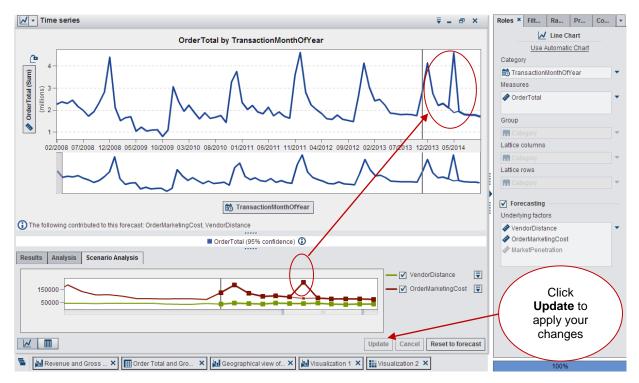


Figure 32

Using scenario analysis you can change the future values of underlying factors to see their impact on the forecast.

The Results tab displays the underlying data that supports this forecast. You can also export this data table as .csv files by right-clicking on the table and selecting **Export Data**.

## SHARING THE RESULTS OF EXPLORATION

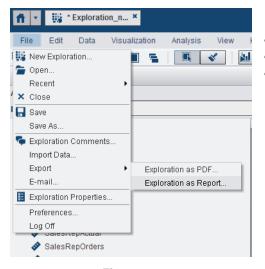


Figure 33:

You can share the exploration by

- Exporting to PDF file
- Exporting as report to the Visual Analytics Designer
- E-mail

# REPORTING USING SAS VISUAL ANALYTICS DESIGNER

On the SAS Visual Analytics home page, click **Create Report** and the designer opens with a new empty report. This interface is very similar to the data exploration with very few differences.

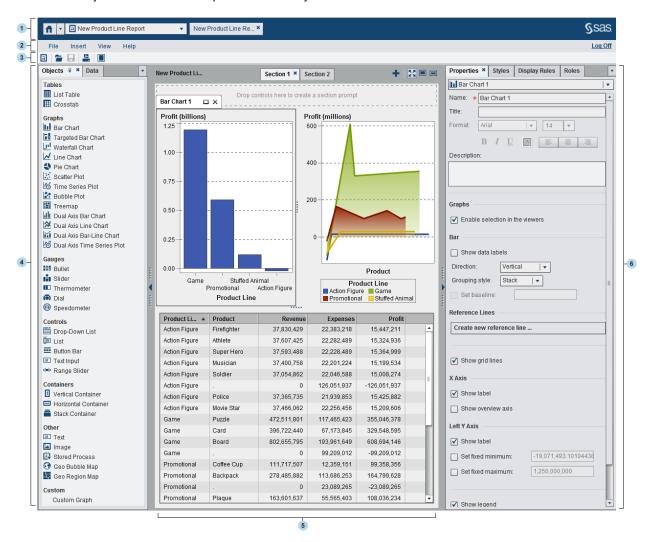


Figure 34:

- 1. The application bar enables you to return to the home page and to access other parts of SAS Visual Analytics and other SAS applications that integrate with the home page.
- The menu bar provides menu items that apply to the entire report or to the currently displayed report section. Actions include creating a new report, adding a new section, inserting new report objects, adding interactions, and launching the viewer without returning to the home page. You can also log off from SAS Visual Analytics.
- 3. The toolbar contains icons that enable you to manage and print your reports. You can click to hide the left and right panes. Or, you can click to display the left and right panes.
- 4. The tabs in the left pane enable you to work with new report objects, data, and imported report objects.
- 5. The canvas is the workspace for building a report. The appearance of the canvas is affected by the report theme.
- 6. A report can have multiple sections, which can be accessed by using the tabs at the top of the canvas. You can change the report view by clicking the icons at the top of the canvas.
- 7. There is an area at the top of the canvas where you can drop filter controls and then categories to create section prompts.
- 8. The tabs in the right pane enable you to work with details about the report and report objects.

## ADDING CONTROLS TO YOUR REPORT SECTION

#### Step1

From the Objects tab, drag and drop two drop down list controls in the top part to create section prompts.

#### Step 2

Assign TransactionYear to one drop down list and facility country/region to the other. The data is loaded along with its frequency for each values. To turn off the frequency on the Roles tab in the right panel select **Remove Frequency**.

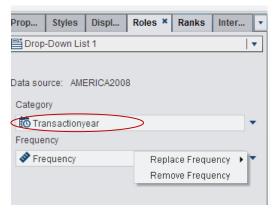


Figure 35

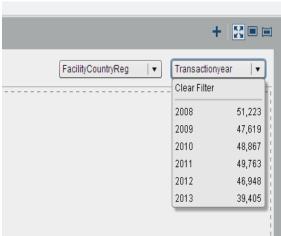


Figure 36

## **VISUALIZNG DATA USING A TREEMAP**

A treemap displays the data as a set of rectangles (called tiles). Each tile represents a category value or a hierarchy node. The size of each tile represents either the frequency of the category or the value of a measure. The color of each tile can indicate the value of an additional measure.

To accomplish this task the steps are as follows:

- From the Objects tab, select and drag the treemap to the report area.
- Under the Roles tab on the right panel, select the product line as the category. Notice that the number and size of the tiles in the maps are changed based on the number and frequency of the product lines
- For size of the tile select order total
- For color of the tile select GrossMarginRatio.

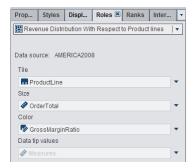


Figure 37

Under the Properties tab, change the name of the visual to "Revenue distribution for product lines."

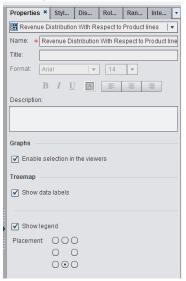


Figure 38

• Frame styling, text styling, and data colors can be changed using the Styles tab in the right panel. The resulting treemap for country selected as US and transaction year as 2009 is shown in Fig.39.

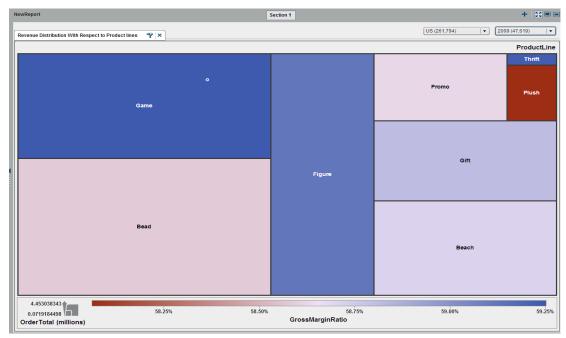


Figure 39

Placing your mouse pointer over each of the rectangles will give you a pop-up window with the values for GrossMarginRatio and OrderTotal for that product line. The figure shows the distribution of GrossMarginRatio and OrderTotal in United States in the year 2009. Now you can see which product groups have the best relative gross margin (in blue) compared to other groups that have same level of revenues (size of the rectangle).

## **ADDING A LINE CHART**

Steps are as follows:

- From the Objects tab, drag and drop a line chart object onto the right side of the report section.
- From the Data tab, drag and drop orderTotal and GrossMargin to the line chart.
- Assign gross margin as a new measure.
- From the Data tab, assign TransactionMonth and year as a new category for the line chart.
- To have the month values in ascending order, right-click the transaction month and year label and select **Sort** and then select **Ascending**.
- Under the Properties tab, select **Overlay Filled** under the Line section and change the name of the line chart to "Revenue Trend".

The resulting visual for Canada at transaction year 2011 is shown in Fig.40.

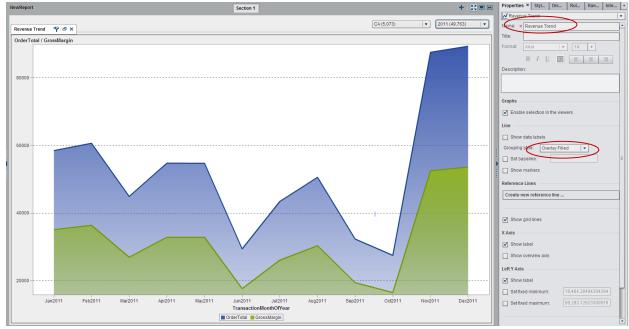


Figure 40

# **ADDING A PIE CHART**

# Steps

- From the Object tab, drag and drop a pie chart object to the report section
- Assign orderTotal as a measure and ProductBrand as the category on the pie chart
- On the Properties tab at the right panel select Show Category Labels and Show Values as Percentage of Total.

The resulting visual is shown below. For the selection for country as US and year as 2009, 58.4% of revenue is from Novelty and 41.6 % of revenue is from Toy.

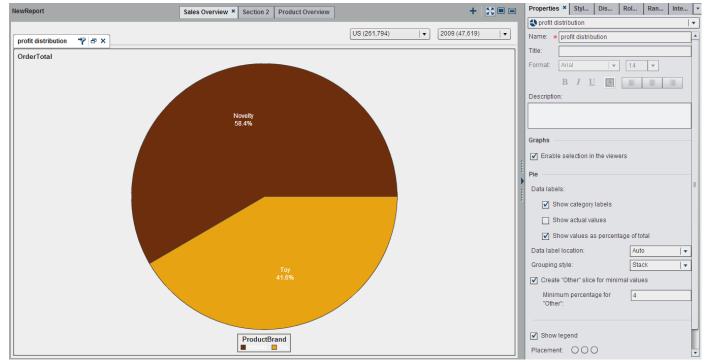


Figure 41

Rename the report section as "Sales overview" and the final report section combining the three objects treemap, line chart, and pie chart is shown in Fig.42.

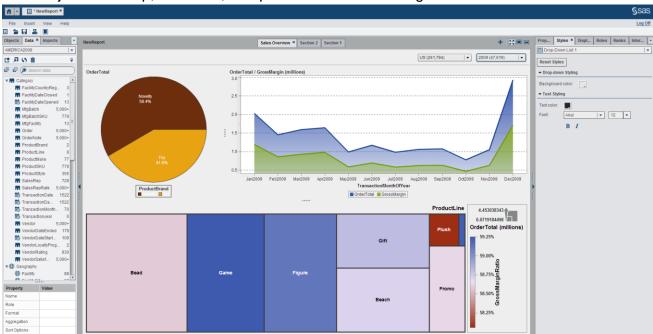


Figure 42

# INTERACTION BETWEEN THE OBJECTS AND DISPLAY RULES

This section will demonstrate two very helpful features of Visual Analytics Designer. The first one is building interactions between the objects in a report section and the second one is adding display rules to the objects. We have used a bar chart and a cross tab for this purpose.

To build a bar chart on a new report section, follow the steps below:

- Add a new section to the above by clicking the button from the top left corner of the report viewer window.
- Select the bar chart from the Objects tab as the new object
- Select Product line as the category, Order Total as the measure, and TransactionYear as the group.
   The resulting bar chart displays the OrderTotal for each product brand grouped by the TrasactionYear. You should be able to identify the trend of the OrderTotal over the years for each of the product line item as shown in Fig.43.

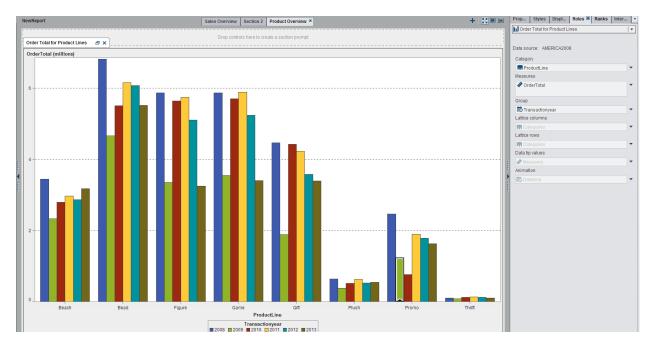


Figure 43

To build a cross tab, follow these steps

- From the Object tab, select a crosstab object.
- Add Facilitycountry/Reg as the column, Product line as the row, and OrderTotal and GrossMargin as the measures and GrossMarginRatio as the measures.
- Select the Styles tab to select the background color and text color for the table. The resulting table is shown below:

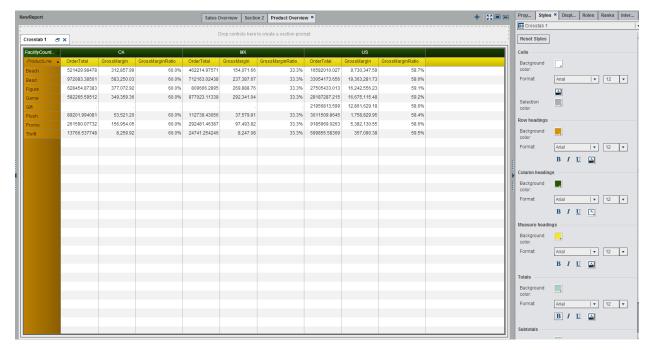


Figure 44:

Now the next step is to build an interaction between the bar chart and the crosstab that enables the user to add a filter or brush interactions between those two objects. Steps to follow are given below:

 Select the interaction view from the right pane under the Interactions tab. The screen will be changed to interaction mode as shown below.



Figure 45:

- Select the interaction type and connect the two objects together as shown above and click Close.
- Now check the interaction that you have defined by selecting a bar on the bar chart. The screen will look something similar to Fig 46

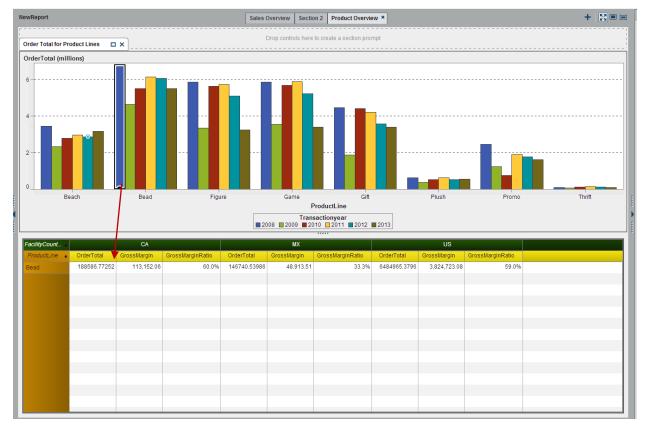


Figure 46:

Notice that the value that is displayed on the table corresponds to the product line "Bead" for transaction year 2008.

The next feature in this section is the addition of display rules to the reporting object. Here we will demonstrate how to add a display rule to the crosstab so that the value of gross margin ratio will be displayed in red if it is less than 50 % and will appear in green if it is greater than or equal to 50 %.

## Step1

Select the crosstab.

## Step2

From the right panel, on the Display Rules tab, select New.



Figure 47:

# Step 3

In the popup window that appears make the selections as shown below and click **OK**.

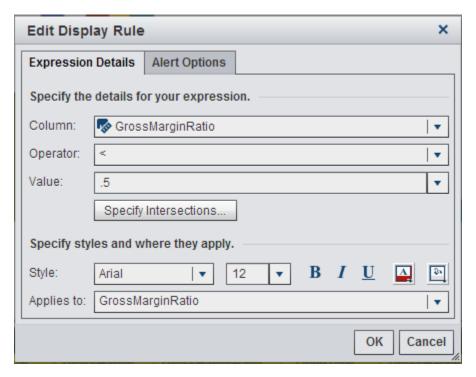


Figure 48:

• Similarly, add another rule by clicking **New** and this time selecting the operator as >= and the color of the text as green. Click **OK**.

The crosstab and the right panel will display the rules as follows:



Figure 49

## CONCLUSION

SAS Visual Analytics provides a powerful environment and toolset for exploring and analyzing your data, and also to create highly interactive and meaningful reports from your data. You can explore any size data quickly with the inmemory analytic capabilities of Visual Analytics. It enables the user to identify trends and patterns in the data and thereby to identify future opportunities.

# **CONTACT INFORMATION**

Your comments and questions are valued and encouraged.

Contact the author at:

Beena Mathew

100 SAS Campus Drive Cary, NC 27513 SAS Institute Inc.

beena.mathew@sas.com

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

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