Self-Service Data Preparation in SAS® Visual Analytics: A Report Designer's Delight

Nicole Ball, Richard Bell, Lynn Matthews, SAS Institute Inc.

ABSTRACT

When you work in SAS® Visual Analytics, you want to complete your reports as quickly and easily as possible. SAS® Visual Analytics 8.3 provides new self-service data manipulation capabilities that significantly increase your control over data sources. This paper examines four of these new capabilities, each of which enables you to work with data more effectively in SAS Visual Analytics. You see how to join multiple SAS® Cloud Analytic Services (CAS) tables in SAS Visual Analytics to incorporate data items from different tables in the same report object. Next, you learn how to define a new data source (an aggregated data source) that can increase efficiency by limiting the number of rows used for an object. You also learn how to define a common filter that can cascade changes to the filter across multiple report objects. And for the final treat, you learn how to capture and reuse data changes as a data view. With this approach, the data view is associated with the CAS table and can be applied anytime that the CAS table is used to create a report. An application administrator can share a data view, so it can be applied by other users who can access the same CAS table. Be prepared to be delighted and to leave this paper a more efficient report designer!

INTRODUCTION

SAS Visual Analytics is a powerful tool that enables you to analyze your data and build reports quickly and easily. Before you can begin using SAS Visual Analytics, you need to ensure that you have the data necessary to perform these tasks. In some instances, you might need to combine, aggregate, modify, or filter your existing tables to create the data needed. In the past, some of this data preparation could only be done outside of SAS Visual Analytics.

Beginning in SAS Visual Analytics 8.3, new self-service data manipulation capabilities enable you to prepare data without having to use another application. These data manipulation capabilities make it easier to quickly get your data in shape without having to rely on data administrators.

This paper discusses four of these new data manipulation capabilities: data source joins, aggregated data sources, data views, and common filters.

The Actions button (⬚) on the Data pane (Figure 1. Actions Button Menu on the Data Pane) enables you to create and use data source joins, aggregated data sources, and data views.
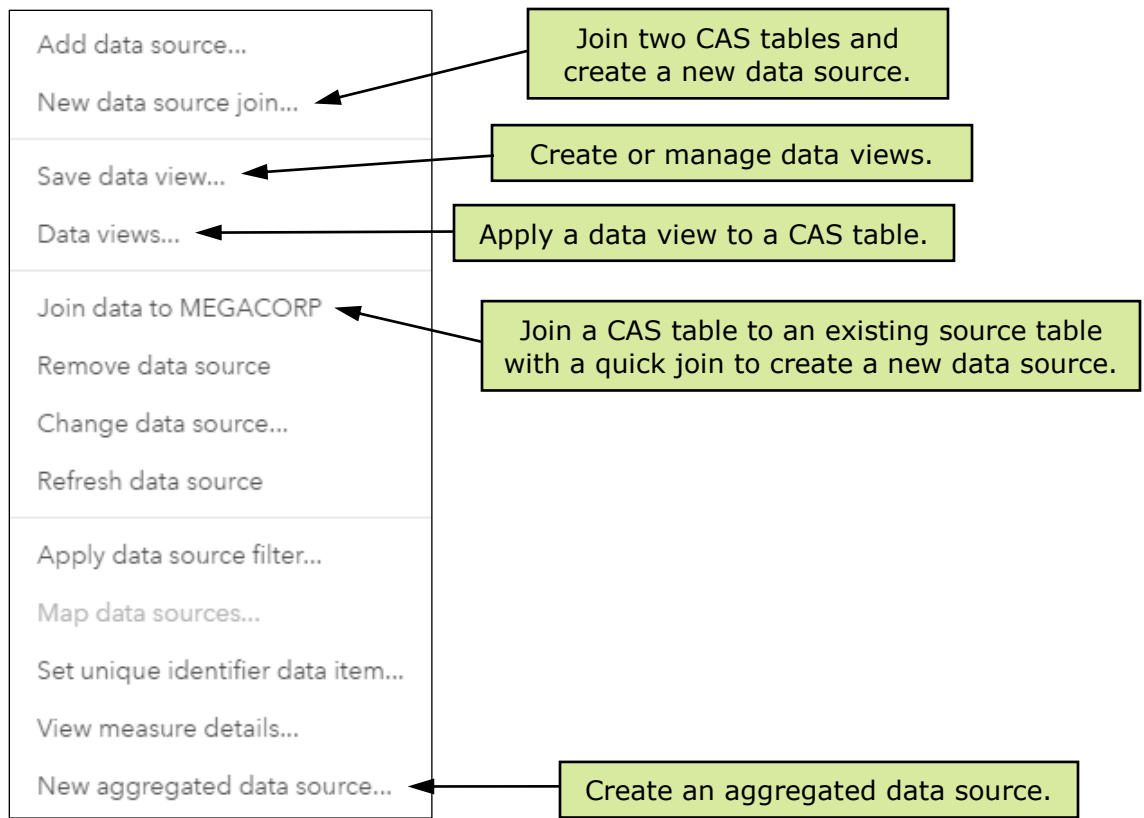
Figure 1. Actions Button Menu on the Data Pane

The common filter, once defined, can be found in a new section in the Data pane (Figure 2. Common Filter Section on the Data Pane).
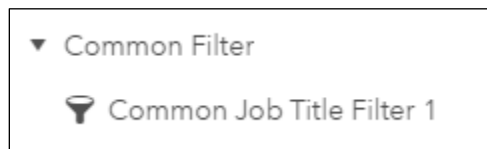

Figure 2. Common Filter Section on the Data Pane

DATA SOURCE JOINS

Before you can begin creating reports in any application, you need to know where your data is located. The data needed to tell your data story might exist in multiple tables and need to be joined before you can use it in your report.

In SAS Visual Analytics, each report object will only accept data items from a single data source. Prior to SAS Visual Analytics 8.3, this presented a challenge. If the data needed for an object was spread among multiple tables, you often needed to request support from a data administrator to prepare the data. This process could take hours, days, or sometime weeks! With the new join capabilities in SAS Visual Analytics 8.3, however, you can join data on the fly within SAS Visual Analytics.

In this example, we want to create a report that shows the average satisfaction of our customers by location. Currently, this data resides in two separate data sources (Figure 3. The Relationship Between CUSTOMERS_TOY and CUSTOMER_LAT_LONG): CUSTOMERS_TOY and CUSTOMERS_LAT_LONG.
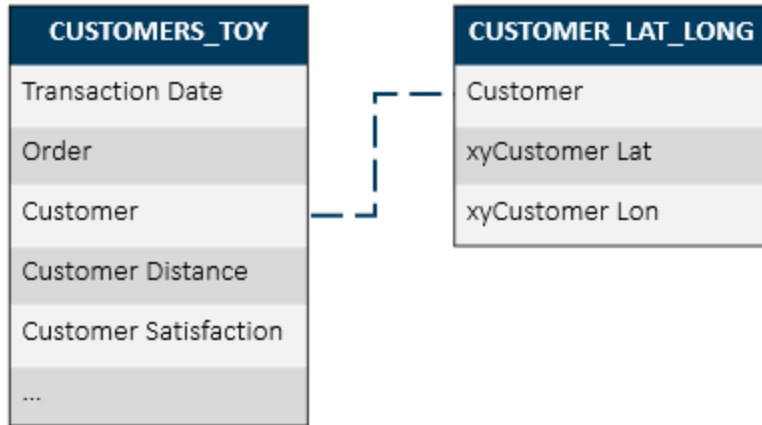
Figure 3. The Relationship Between CUSTOMERS_TOY and CUSTOMER_LAT_LONG

CUSTOMERS_TOY contains details about our customers, their orders, and their satisfaction. CUSTOMERS_LAT_LONG contains the latitude and longitude location of each customer. To create the requested report, we will need to join the two tables together by the ID of the customer (labeled Customer in both data sources).

You can join data within SAS Visual Analytics by either creating a new data source join (new join) or by joining data to the source table (quick join). Both options can be accessed through the Actions button ( ) on the Data pane (Figure 4. Join Options on the Actions Button Menu).
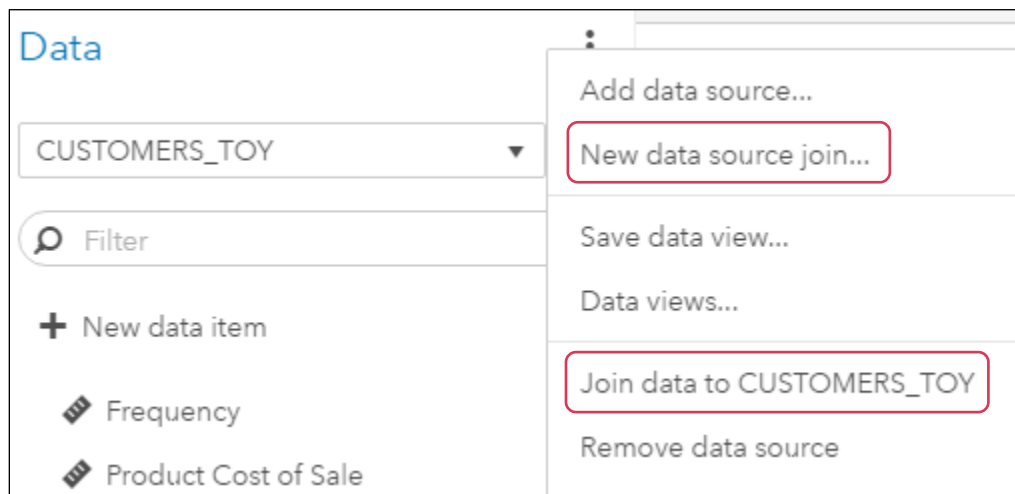

Figure 4. Join Options on the Actions Button Menu

To create a new data source join (new join), first add one of the data sources to the report. Then select (**Actions**) ⇨ **New data source join** on the Data pane. This opens the New Data Source Join window (Figure 5. New Data Source Join Window) where you can specify the details of the join: the name, the join type, the data sources to join, the join conditions, and the columns to include in the new table.

Figure 5. New Data Source Join Window

The new data source join gives you full control over specifying the properties for the join.

To join data to the source table (quick join), first add one of the data sources (CUSTOMERS_TOY) to the report. Then select 🖳 (**Actions**) ⇨ **Join data to CUSTOMERS_TOY** on the Data pane. This opens the Open Data Source window where you can select the joining table (CUSTOMER_LAT_LONG). A quick join is created between the two tables. The quick join uses a default name (Data_Source_Join), the default join type (left join), joins on all categories with matching names or labels, and selects all columns (minus one of the join condition columns) from both tables. If no category data items have matching names or labels, measures will be used instead.

The properties for both the new join and the quick join can be modified by selecting 🖳 (**Actions**) ⇨ **Edit data source join** on the Data pane (Figure 6. Edit Data Source Join Option on the Actions Button Menu).
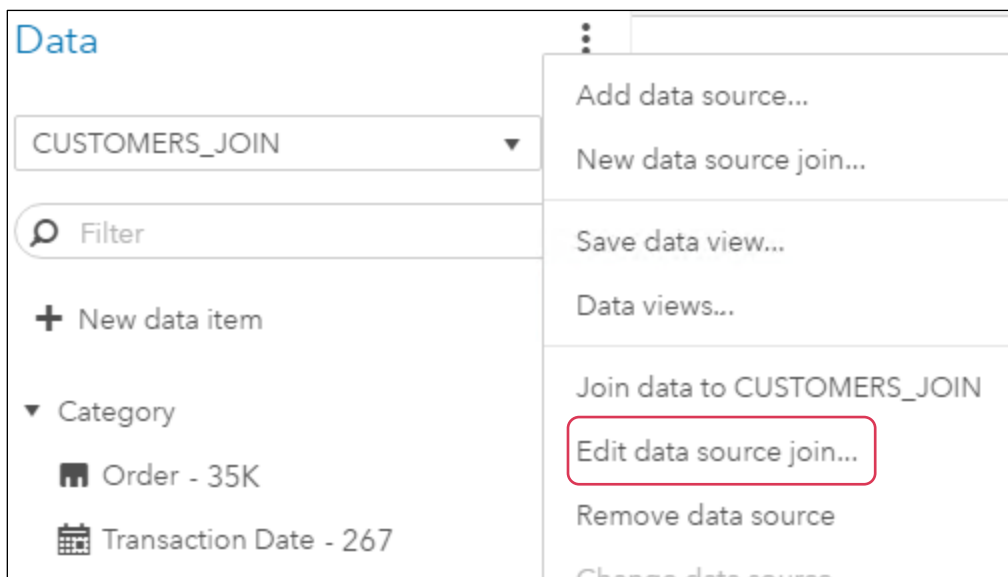

Figure 6. Edit Data Source Join Option on the Actions Button Menu

Both the new join and the quick join create a temporary table when executed, which is added as a new data source to the Data pane in SAS Visual Analytics. This temporary table is cleaned up when no longer needed. It is available to use for the current report, but it is not available for other users. An administrator can view details about the new data source join in SAS® Environment Manager, but the table uses a temporary name (_VA_*SOURCE_TABLE_NAME_unique-string*).

Once the tables are joined, you can create the requested map of average satisfaction by customer location (Figure 7. Geographic Map of Average Customer Satisfaction by Location).

Figure 7. Geographic Map of Average Customer Satisfaction by Location

Before using a data source join, keep the following points in mind:

- Data source joins and data source filters cannot be used together. A table with a data source filter cannot be one of the join tables. A data source filter cannot be applied to the joined table.

- The join tables cannot contain calculated data items and custom categories. However, you can add calculated items and custom categories to the joined table.

- For matching purposes, data values are evaluated using the formatted value. This could mean that two identical data items might not match if the applied formats are different. However, you can change the format in one of the data sources using the Data pane to work around this issue

AGGREGATED DATA SOURCES

When you add data to report objects in SAS Visual Analytics, the data is often aggregated for display. This aggregation of data takes time and resources. Some reports might contain multiple objects that display the same level of data.

Beginning in SAS Visual Analytics 8.3, you can create aggregated data sources. This increases the efficiency of your reports because the number of rows used for each object is limited. Aggregated data sources are typically smaller versions of the original table that contain totals (or aggregations) for a subset of columns. Applying the aggregated data source to multiple objects could result in faster display of the report and fewer resources utilized.

In this example, we have a report that displays the same level of data for multiple report objects. The Summary page (Figure 8. Summary Page) displays a bar chart that shows profit by product line and a treemap that shows profit and number of products by product group.
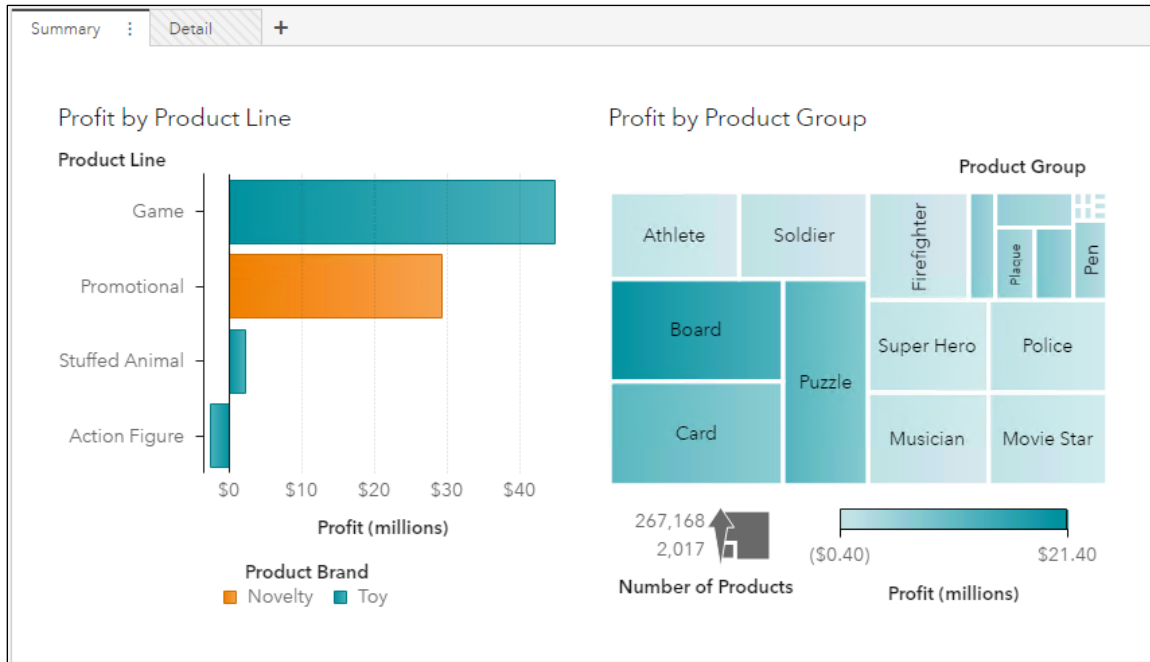

Figure 8. Summary Page

The Detail page (Figure 9. Detail Page) displays a list table with the profit and number of products for each product type.

Figure 9. Detail Page

Instead of querying the detail table each time the report is displayed, we want to use an aggregated data source for these objects to try to reduce the display time and the resources used to render this report.

You can create an aggregated data source by selecting ⊞ (**Actions**) ⇨ **New aggregated data source** on the Data pane. This opens the New Aggregated Data Source window (Figure 10. New Aggregated Data Source Window) where you can specify the details for the aggregation: the name, the columns, and any filters. After selections are made, you can see a preview of the result table.

Figure 10. New Aggregated Data Source Window

The list of available data items includes all data items except hierarchies, geography items, spline effects, partitions, scoped calculated items, interaction effects, calculations that contain a suppressed or time-period calculation, and calculations that use the AggregateCells operator.

7

When you create an aggregated data source, you want to select columns with small distinct counts. The fewer distinct values, the smaller the aggregated data source, and the bigger the time and resource savings. In this instance, we will select Product Brand, Product Line, Product Group, and Product Type (Figure 11. Selected Category Data Items).
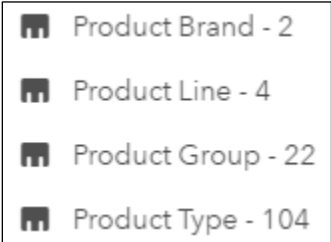

Figure 11. Selected Category Data Items

The aggregated table will contain one row for each distinct crossing of these four categories. Any measures that are specified (Profit and Number of Products) are totaled for the aggregated table (Figure 12. Selected Measure Data Items).
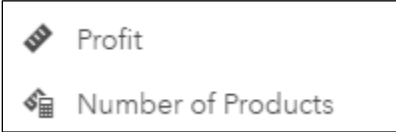

Figure 12. Selected Measure Data Items

The aggregated data source appears as another data source in the report (Figure 13. Aggregated Data Source) and can be used to populate report objects.
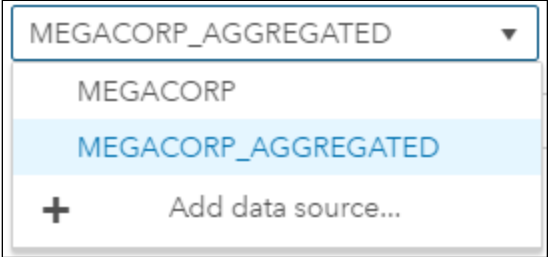
Figure 13. Aggregated Data Source

The detail table (MEGACORP) contains 46 columns and 2,229,087 rows of data, but the aggregated table (MEGACORP_AGGREGATED) contains only 6 columns and 134 rows of data. You can add new data items to the aggregated table (for example, hierarchies and custom categories) and modify existing data items just like any other table.

Next, you need to modify the objects to use the data in the aggregated table. Right-click an object in the canvas and select **Remove** ⇨ **All role assignments**. Verify that the aggregated table is selected in the Data pane and add the data to the report object using the Roles pane.

After the report has been modified to use the aggregated table instead of the detail table, you'll need to do some benchmarking to ensure that your report performance has improved. You can access performance statistics for the report by clicking Ctrl+Alt+P and selecting Data in the drop-down list (Figure 14. SAS Visual Analytics Diagnostics Window).

Figure 14. SAS Visual Analytics Diagnostics Window: Performance

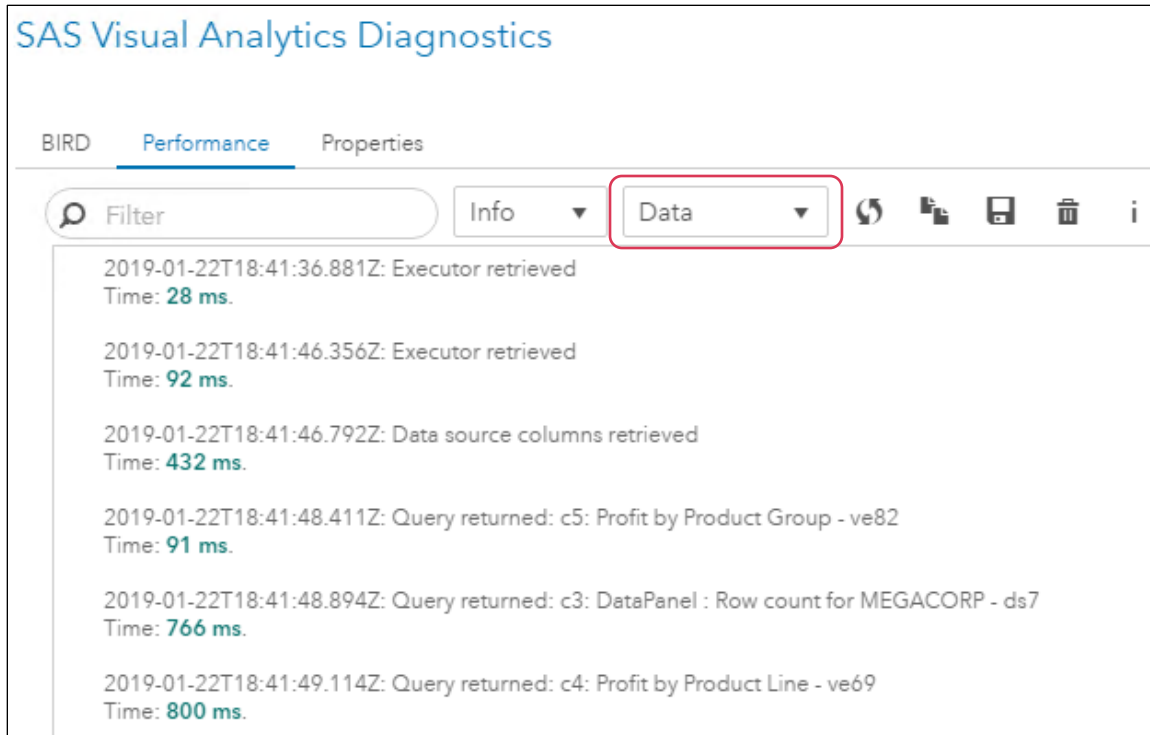Be aware that using an aggregated data source does not guarantee increased performance and the improvements (if any) will vary. Results will depend on the number of objects that use the aggregated table, the filters, the calculations, and the level of detail displayed in the object. In addition, creating the aggregated data source utilizes resources and might offset any benefits you might receive. Benchmarking will help you determine if the creation and use of the aggregated table is beneficial for the report.

GROUPING AGGREGATED MEASURES

Grouping aggregated measures is another use for aggregated data sources.

In this example, we have a table (MEGACORP) that contains details about our products. We have already added an aggregated measure (Number of Products) to the table that counts the distinct product IDs (Figure 15. Expression for Number of Products).

Figure 15. Expression for Number of Products

For our report, we would like to categorize the number of products for each product group into different ranges:

- *Low*: The number of products is less than or equal to 100,000.

- *Medium*: The number of products is between 100,000 and 200,000.

- *High*: The number of products is greater than or equal to 200,000.

Figure 16. Number of Products by Product Group shows the current number of products for each product group.

9

Product Group ▲	Number of Products
.	1
Athlete	166,839
Backpack	41,567
Bear	2,074
Big Cats	2,106
Board	262,325
Card	267,168
Cat	2,054
Coffee Cup	41,186
Dog	2,054
Elephant	2,040
Firefighter	161,028

Figure 16. Number of Products by Product Group

If Number of Products was a measure rather than an aggregated measure, you could create a custom category to group the ranges or create a calculated data item to create the groups. However, it is not possible in SAS Visual Analytics to create a custom category from an aggregated measure; only categories and measures can be used (Figure 17. Only Categories and Measures Are Available for Custom Categories).

Based on:

Product Brand - 2 ▼

- Product Brand - 2
- Product Description - 105
- Product Group - 23
- Product ID - 2.2M
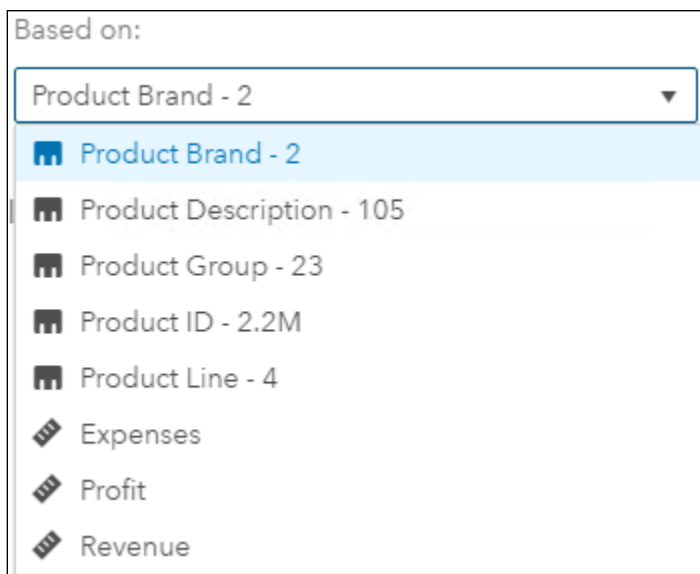- Product Line - 4
- Expenses
- Profit
- Revenue

Figure 17. Only Categories and Measures Are Available for Custom Categories

You also cannot create a calculated column from an aggregated measure (Figure 18. Aggregated Measures Cannot Be Used in a Calculated Column).

Figure 18. Aggregated Measures Cannot Be Used in a Calculated Column

You can create calculated columns, however, with aggregated data sources. First, you need to create a small aggregated table of Product Group and Number of Products, which has 23 rows and 2 columns (Figure 19. New Aggregated Data Source MEGACORP_PRODNO).
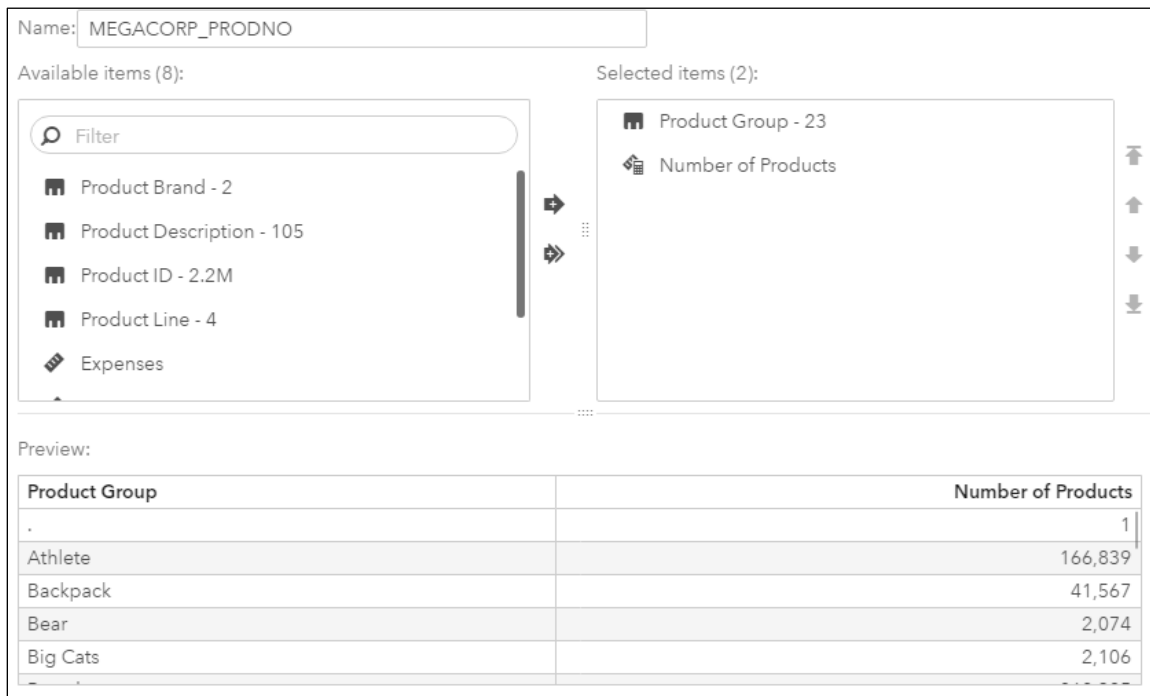

Figure 19. New Aggregated Data Source MEGACORP_PRODNO

In the aggregated table (MEGACORP_PRODNO), Number of Products appears as a measure, instead of an aggregated measure (Figure 20. Data Pane for MEGACORP_PRODNO). This means that you can perform all types of calculations on this data item including custom categories and calculated columns.

11

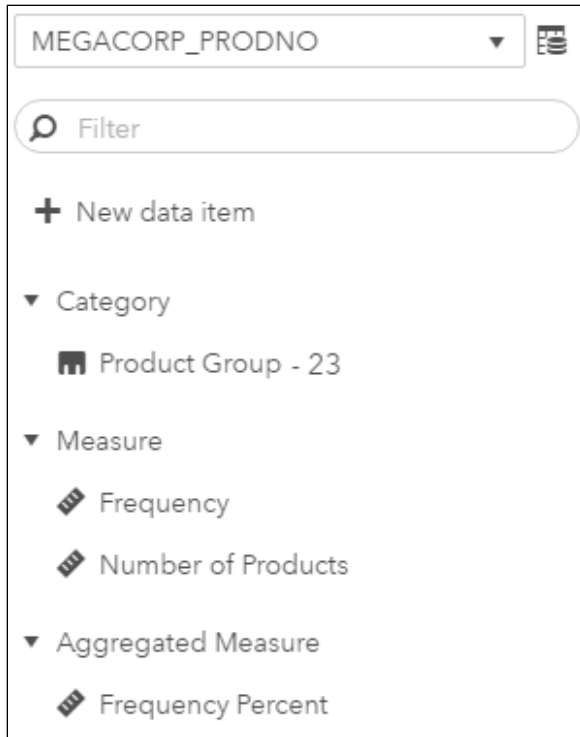**Figure 20. Data Pane for MEGACORP_PRODNO Data Table**

For our example, we can create a custom category that categorizes the number of Products (Figure 21. Custom Category Properties to Categorize Products).
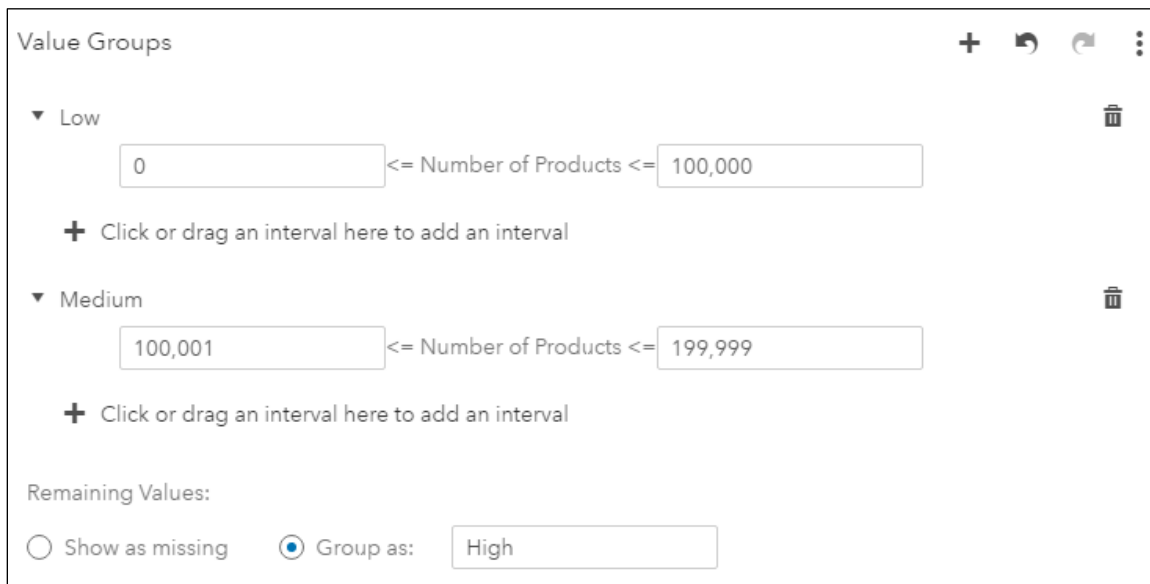
Figure 21. Custom Category Properties to Categorize Products

We can also create a calculated data item that categorizes the number of products (Figure 22. Calculated Item Expression to Categorize Products).

Figure 22. Calculated Item Expression to Categorize Products

These new data items can be used in any report object to show categorized product groups or the number of product groups that contain low, medium, or high number of products (Figure 23. List Table with Custom Category and Calculated Item and Figure 24. Bar Chart of Calculated Item).

Product Group ▲	Number of Products	Categorize Products (custom category)	Categorize Products (calculated item)
.	1	Low	Low
Athlete	166,839	Medium	Medium
Backpack	41,567	Low	Low
Bear	2,074	Low	Low
Big Cats	2,106	Low	Low
Board	262,325	High	High
Card	267,168	High	High
Cat	2,054	Low	Low
Coffee Cup	41,186	Low	Low
Dog	2,054	Low	Low
Elephant	2,040	Low	Low
Firefighter	161,028	Medium	Medium
Horse	2,017	Low	Low
iPhone Cover	42,258	Low	Low

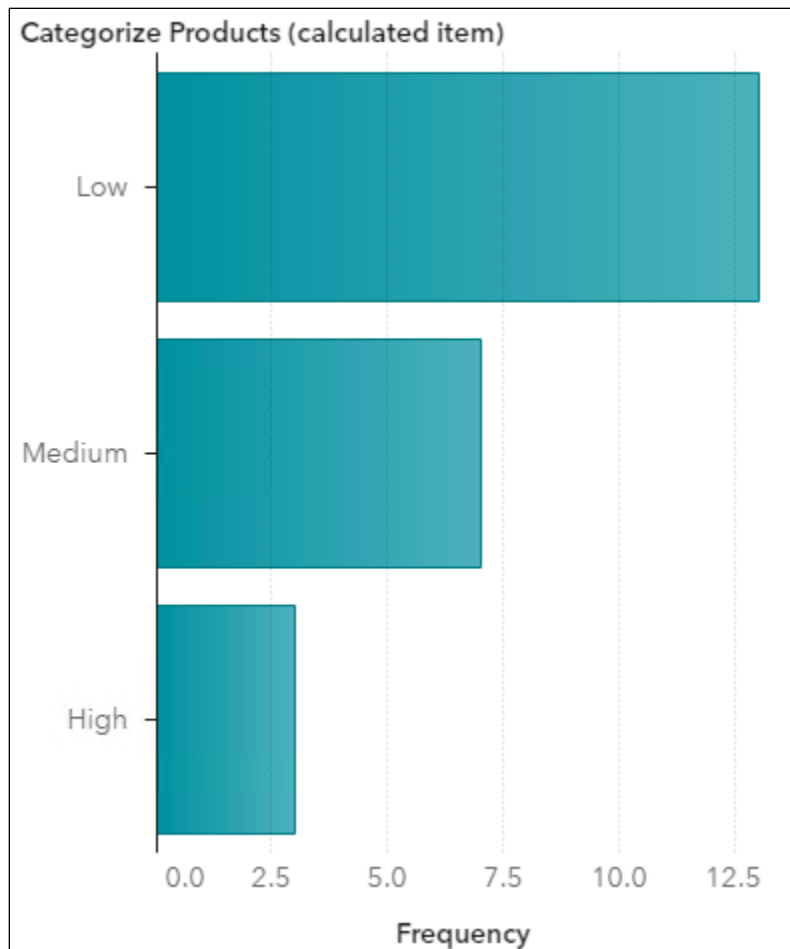**Figure 23. List Table with Custom Category and Calculated Item**

Figure 24. Bar Chart of Calculated Item

Before using an aggregated data source, keep the following points in mind:

- The aggregated data source is generated when the report is opened and is only available for the current report.

- The rows of the aggregated data source are summarized for the categories selected. Fewer selected categories results in fewer rows.

- The rows of the aggregated data source are based on the aggregation in the source table. For example, if Customer Satisfaction has an aggregation of Average in the source table, the rows of the aggregated data source will display average customer satisfaction for the selected categories. If the aggregation in the source table is modified, then the aggregated data source is regenerated based on the new aggregation.

- Two aggregated data sources that do not have any data source filters can be joined using a data source join.

- When creating an aggregated data source using a date or datetime data item, it is a good idea to ensure that the formats for these items include the level of detail you need. The data in the aggregated data source is stored using the format specified at the time the aggregated data source was created. For example, if Transaction Date has a Month format but you later discover you also need to see years, you cannot change the format in the aggregated data source to add this information.

- If you create an aggregated data source with a data item that uses a user-defined format, the format

15

is no longer displayed as user-defined in the Data pane. The values are displayed using the format, but the format of the data item in the aggregated data source cannot be removed or modified.

- When Frequency and Frequency Percent are included in an aggregated data source, they are automatically renamed to Aggregated Frequency and Aggregated Frequency Percent. They are both listed in the Measure group on the Data pane.

DATA VIEWS

When you create reports, sometimes data changes are necessary. Some of these changes can be made to the CAS table (by building plans in SAS® Data Studio or by writing code). However, some analysts do not have access to data prep tools or the changes need to be made directly in SAS Visual Analytics (for example, creating geography data items, hierarchies, aggregated measures, modifying aggregations).

You can make changes to the data in SAS Visual Analytics, but those changes are saved with the report and not written back to the CAS table. Prior to SAS Visual Analytics 8.3, this created a problem. One table could require numerous changes every time it was used. Instead of spending valuable time working with the data, a template report could be created that included all the necessary data changes. That template could then be used as a starting point for any new reports built using the table. While a clever solution, the template was just one more thing that needed to be monitored. Documentation about the template needed to be created and updated, the template needed to be saved where it could be accessed, and assurances needed to be made so that someone didn't accidentally overwrite the template.

With the new data view capabilities in SAS Visual Analytics 8.3, you can create a data view that is saved with (and can be automatically applied to) a CAS table. The data view captures the following data settings for your report:

- data item properties (names, formats, classifications, and aggregations)

- data source filters

- geography data items

- hierarchies

- calculated items

- aggregated measures

- custom categories

- duplicate data items

- show and hide statuses for data items

- unique row identifiers

In this example, we have a table (PRODUCTS_CLEAN) that has already been cleaned and prepped using SAS Data Studio for the purposes of reporting. However, the following modifications still need to be made to this table:

- Add a data source filter for the United States.

- Modify the data item properties for names and aggregations (Table 1. Modifications to Data Item).

- Create the geography data items State Name and Postal code.

- Create the geographic hierarchy State Name ⇨ Postal code.

- Create the product hierarchy Product Line ⇨ Product Category ⇨ Product Group ⇨ Product Name.

- Create the calculated item Profit = Total Revenue – Unit Cost * Quantity.

- Create the derived data item Number of Products.

- Hide unneeded data items.

Table 1. Modifications to Data Items

Data Item	Modification
Cost	Rename: Unit Cost
Discount in percent of Total Normal Retail Price	Aggregation: Average Rename: Average Discount
Quantity Ordered	Rename: Quantity
Retail Price	Rename: Total Revenue

Figure 25. Data Pane for PRODUCTS_CLEAN shows the Data pane for the PRODUCTS_CLEAN table after the data modifications are completed.
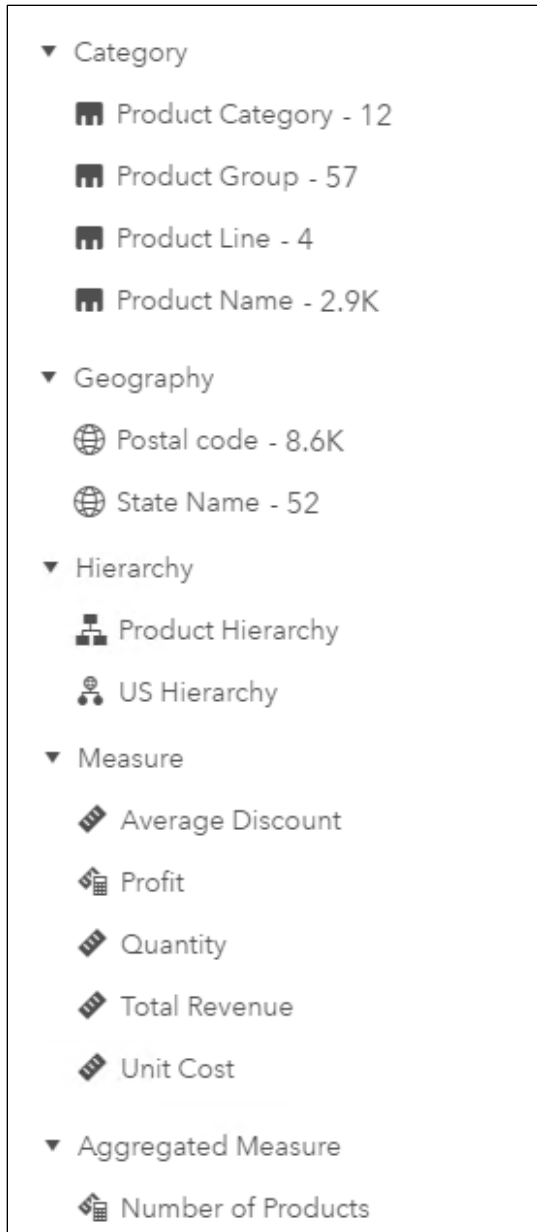
Figure 25. Data Pane for PRODUCTS_CLEAN Data Table

After making the necessary changes to the data in SAS Visual Analytics, you can create a data view by selecting ▤ (**Actions**) ⇨ **Save data view**. This opens the Save Data View window (Figure 26. Save Data View Window) where you can specify details for the view: the name, a description, whether to set it as default, and/or whether to set it as shared.
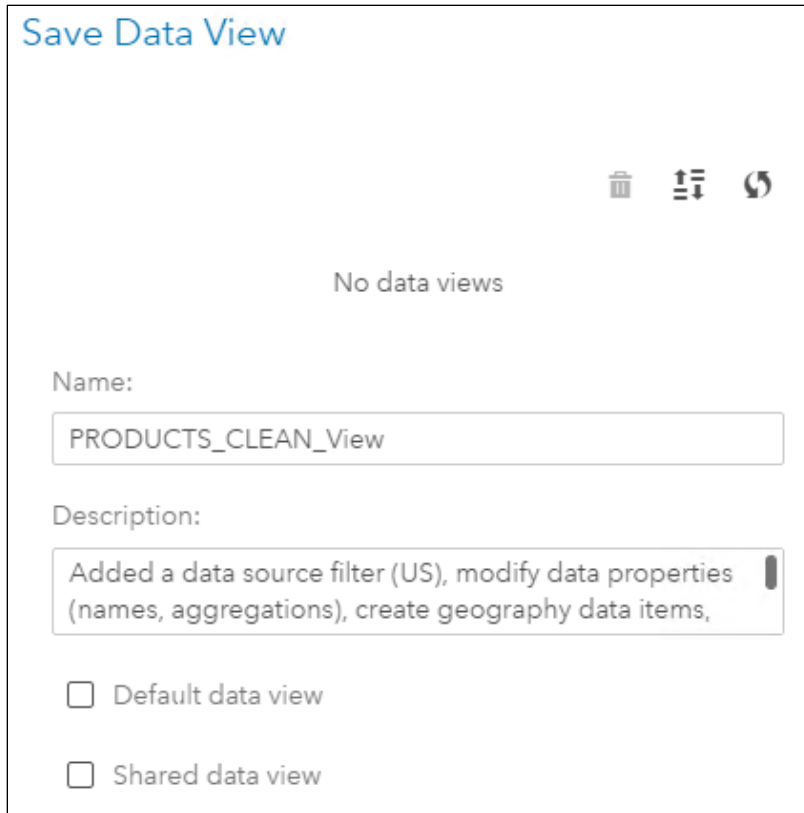
Figure 26. Save Data View Window

Note: The **Shared data view** option is available only to users who are members of the Application Administrators group.

If a data view is not saved as the default, the view is not applied automatically when the table is added to a report. The view can be applied to the table by selecting ▦ (**Actions**) ⇨ **Data views**. The Data Views window (Figure 27. Data Views Window) lists all views associated with the table.

Figure 27. Data Views Window

When the data view is applied, any differences with the existing data will be applied additively. This means that any data items that exist in the current report, but do not exist in the data view, are retained. Any conflicts between data in the current report and the data view are resolved by creating duplicate items.

In this example (Figure 28.), the original table contained a column called *Discount in percent of Normal Total Retail Price*. For the data view, this column was renamed (*Average Discount*) and the aggregation was changed to Average. When the data view is applied to

the table, the original column remains and a new column with the data changes is added to the table.
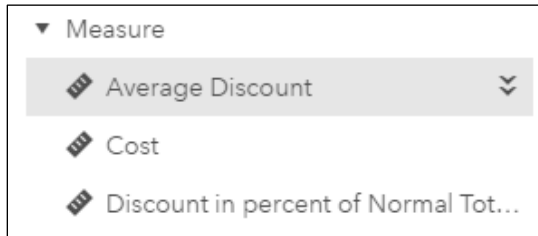

Figure 28. Data Items Duplicated When View is Applied

If a data view is saved as the default, the view is automatically applied when the table is added to the report. In this case, when the table is added only the data items for the view appear.

Multiple data views can be created for a single table, but only one can be designated as the default. If multiple data views are available, the default view is marked in the Data Views window (Figure 29. Multiple Data Views).
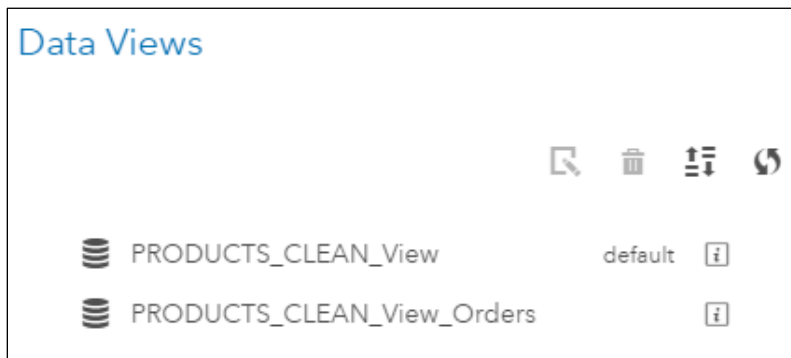

Figure 29. Multiple Data Views

A data view can only be accessed, applied, and modified by the analyst that created it. To enable others to use the data view, an application administrator can save the data view as a shared view.

If a data view is saved as a shared view, the view can be accessed, applied, and modified by any users that can access the CAS table. Application administrators can also specify a default data view for all users by selecting ⋮ (**Menu**) ⇨ **Edit administration settings** in SAS Visual Analytics.

In the Default Data View section of the Administration Settings window (Figure 30. Administration Settings: Default Data View), default data views can be defined for each table that contains shared data views.
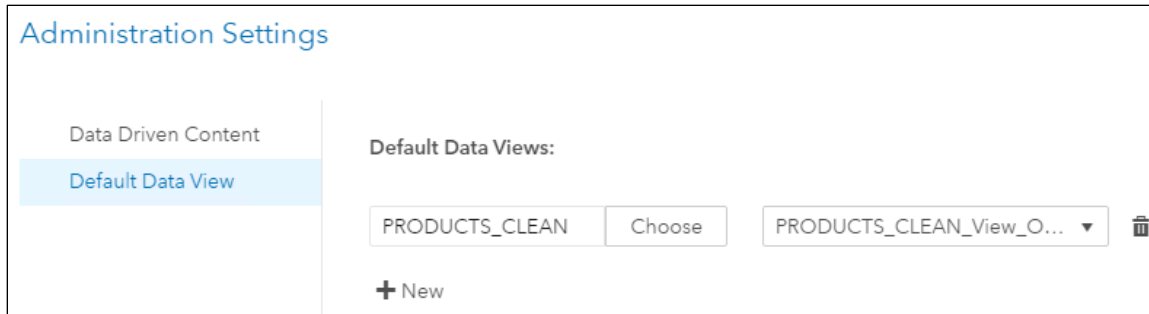
Figure 30. Administration Settings: Default Data View

If a user has not specified a default data view for the table, the administrator default view will be applied automatically. Users can override this setting by selecting another view as their default.

Before using a data view, keep the following points in mind:

- Each data view is specific to one table. You cannot apply a data view to other tables.

- Data views are saved separately from reports. If you create a data view in one report, you can apply it to other reports.

- A data view acts as a template for data source settings. Updating a data view does not automatically update existing reports that use the view.

- An administrator can set a default data view for a table. You can also set a personal default. Your personal default data view will override any administrator defaults.

- Default data views are applied automatically when the table is added to a report.

- Data views cannot be created for aggregated data sources that contain derived data items.

COMMON FILTERS

In SAS Visual Analytics, you can create two different types of filters: filters that can be modified by only the report designer (fixed filters), and filters that can also be modified by the report viewer (interactive filters). Fixed filters can be data source filters or report object filters. Data source filters are applied to all objects in the report that use the data source. Report object filters, on the other hand, are applied to a single report object.

In some cases, you might want to apply the same filter to multiple report objects (but not to the entire data source). Prior to SAS Visual Analytics 8.3, this meant that you had to create the filter for each object separately. If the filter was very complex, this could waste a lot of time and leave room for mistakes. In addition, each object filter was independent of the others. Each time a change was required, you would need to modify each filter one by one. With the new common filter capabilities in SAS Visual Analytics 8.3, the same filter can be applied to multiple report objects and modified for all objects in one location.

In this example (Figure 31. Employees Report), we have an existing report that shows details about our employees.
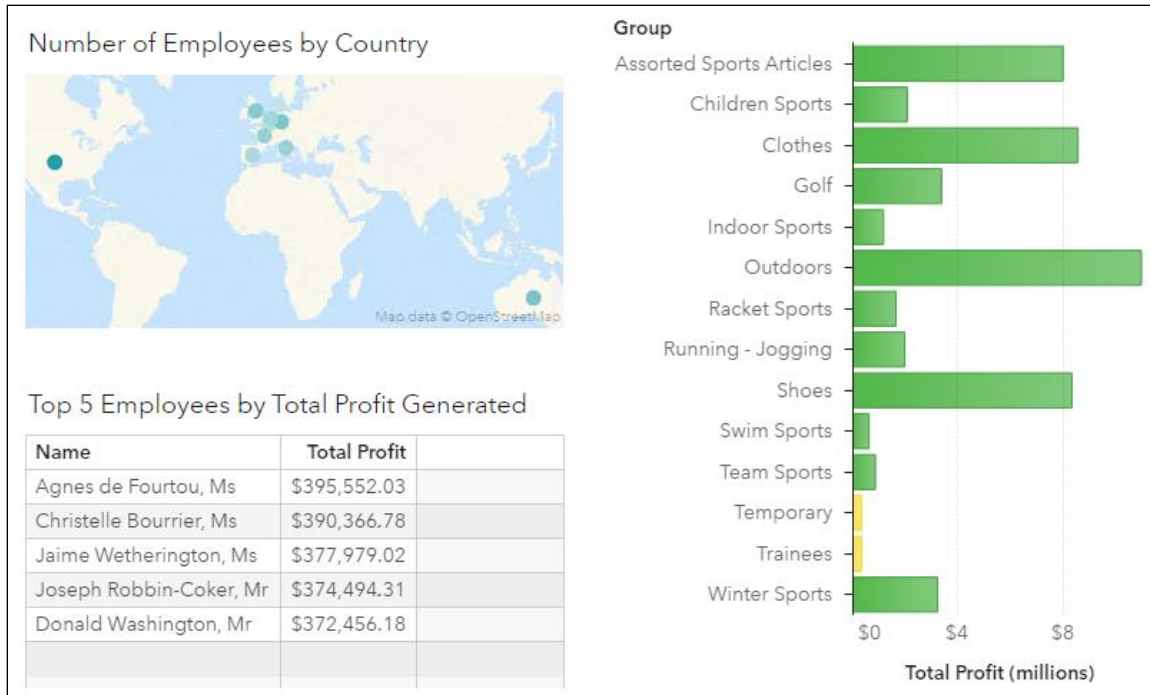
Figure 31. Employees Report

The report contains a geo map that shows the number of employees by country, a list table with the top five employees by total profit generated, and a bar chart that shows the total profit by group.

We want to add a fixed filter to only the geo map and the list table to analyze the differences for different job titles. We can do this with a common filter.

After you add a report object filter to one object (the geo map), you can create a common filter from the Filters pane by selecting ⋮ (**Options**) ⇨ **Change to common filter** (Figure 32. Options Menu on Filters Pane).
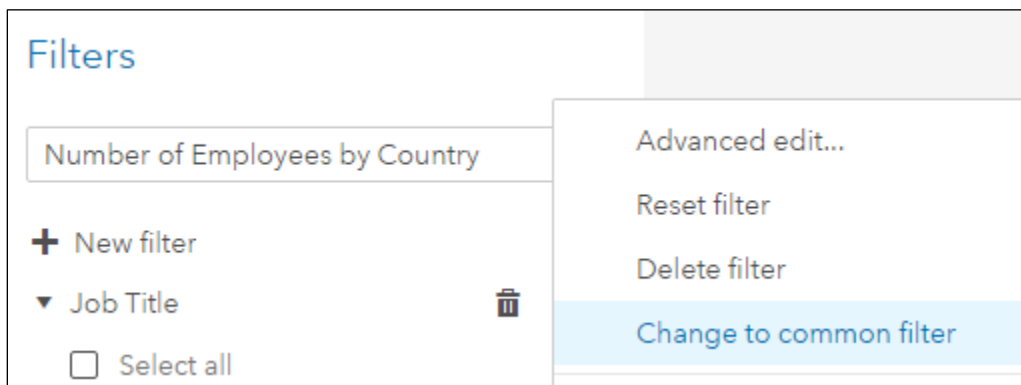

Figure 32. Options Menu on Filters Pane

This filter currently selects all job titles except Sales Rep. I, Sales Rep. II, Sales Rep. III, and Sales Rep. IV.

The common filter is added to the Data pane under a new section, Common Filter (Figure 33. Common Filter Section on Data Pane). The common filter can be renamed, if necessary.

Figure 33. Common Filter Section on Data Pane

Once the common filter has been created, it can be applied to any other object in the report by simply dragging the common filter to the object (Figure 34. Add Common Filter to List Table).

Figure 34. Add Common Filter to List Table

Another way to add the common filter to an object is to select the object in the canvas, right-click the Common Filter in the Data pane, and select **Add to selected object**. In addition, the common filter can be added via the Filters pane by selecting **New filter** (Figure 35. Add Common Filter Using the Filters Pane).
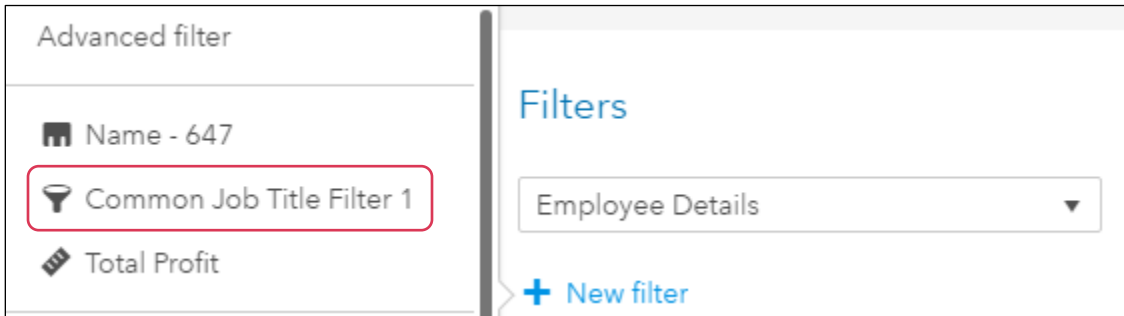

Figure 35. Add Common Filter Using the Filters Pane

After the common filter has been added to the necessary objects, the selected values can be edited by selecting any of the filtered objects and modifying the filter values in the Filters pane. The changes are cascaded to any objects that use the common filter.

In this example (Figure 36. Modify the Common Filter), we modified the common filter to only select the Sales Rep. I job title.
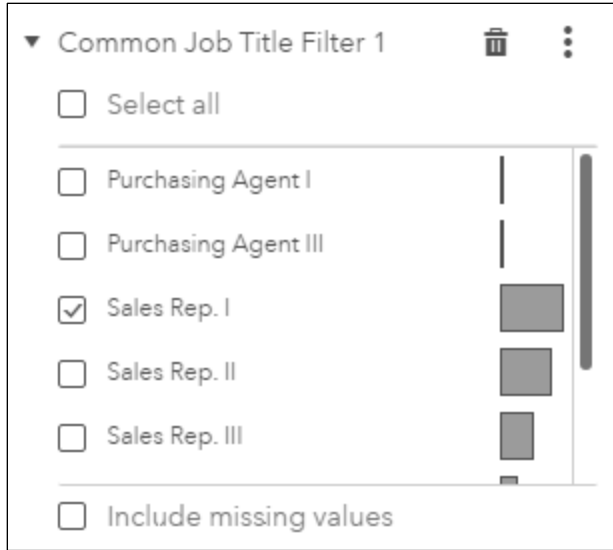
Figure 36. Modify the Common Filter

Both the geo map and the list table are updated to show information about that job title (Figure 37. Geo Map and List Table Filtered by Common Filter).
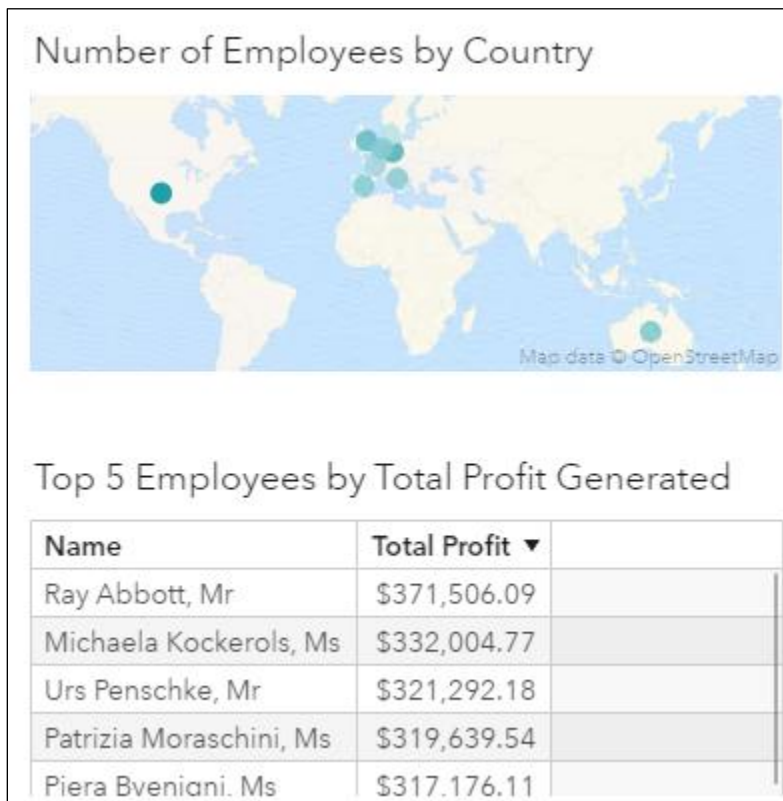

Top 5 Employees by Total Profit Generated

Name	Total Profit ▼	
Ray Abbott, Mr	$371,506.09	
Michaela Kockerols, Ms	$332,004.77	
Urs Penschke, Mr	$321,292.18	
Patrizia Moraschini, Ms	$319,639.54	
Piera Byeniani, Ms	$317,176.11	

Figure 37. Geo Map and List Table Filtered by Common Filter

The common filter can be deleted from any object using the Filters pane. Simply select the object in the canvas, and in the Filters pane, click 🗑 (**Delete**) next to the common filter. This only deletes the common filter from the object, not from the report. To delete the

common filter from the report, right-click the common filter in the Data pane and select **Delete**.

The common filter can also be converted to a report object filter by selecting ⋮ (**Options**) ⇨ **Copy to object-specific filter** on the Filters pane. Any changes made to the object-specific filter will only affect that object and will not modify the common filter.

Before using a common filter, keep the following points in mind:

- Path filters for path analysis objects cannot be converted to common filters.

- Common filters are specific to a report and cannot be shared between reports.

CONCLUSION

The new self-service data preparation capabilities in SAS Visual Analytics 8.3 enable you to increase your efficiency by enabling you to prepare and modify your data without having to rely on a data administrator. Data source joins give you the ability join data on the fly, aggregated data sources can help improve the efficiency and reduce the resources used to render a report, and data views can be created to minimize the time spent modifying and creating data items for the report. In addition, common filters enable you to easily apply and modify a filter that is applied to multiple report objects.

REFERENCES

SAS Institute Inc. 2018. SAS® Visual Analytics 8.3: Working with Report Data. Cary, NC: SAS Institute Inc. Available at https://go.documentation.sas.com/?cdcId=vacdc&cdcVersion=8.3&docsetId=vareportdata&docsetTarget=titlepage.htm&locale=en.

ACKNOWLEDGMENTS

We would like to thank our manager, Stacey Syphus, for her support of this work. In addition, we appreciate Melissa Oswald for taking time to review the paper in its first draft and Rick Cornell for reviewing and editing the abstract.

CONTACT INFORMATION

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

Nicole Ball
SAS Institute, Inc.
Nicole.Ball@sas.com

Richard Bell
SAS Institute, Inc.
Richard.Bell@sas.com

Lynn Matthews
SAS Institute, Inc.
Lynn.Matthews@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.