

Paper 051-2013

Self-Service Data Management: SAS® Visual Data Builder

Sam Atassi, SAS Institute Inc., Chicago, IL

Malcolm Alexander, SAS Institute Inc., Cary, NC

ABSTRACT

Successful data preparation is the key to extracting meaningful knowledge from data. SAS® Visual Data Builder allows you to access data from enterprise sources and transform it for use in business intelligence, data visualization, and data mining tasks. This paper discusses self-service data management techniques available using SAS® Visual Data Builder as well as its unique features enabling it to load data into SAS® LASR™ Analytic Server.

INTRODUCTION

Organizations today strive to find innovative ways to analyze, evaluate, interpret, and infer answers from oceans of data. IT, increasingly, struggles to keep up with the demands for data structuring and preparation required to enable these high-value activities. With SAS® Visual Analytics, SAS is making analytics approachable, providing business users with insights into their data with easy-to-create dashboards and graphs. An often underappreciated part of the process is the effort to prepare data for any number of follow-up tasks such as predictive analytics, forecasting, or query and reporting.

SAS Visual Data Builder, a component of SAS Visual Analytics, allows you to query data from various databases, structure it for analytics and reporting, and then load into any SAS library that is available. In particular, SAS Visual Data Builder makes it easy to load data into the SAS LASR Analytic Server for interactive exploration or report creation, or for high-performance analytics with SAS® High-Performance Analytics.

In addition, SAS Visual Analytics makes big data accessible, providing the power to analyze large volumes of data. However, extraction and structuring of big data is more challenging. SAS Visual Data Builder has capabilities for scheduling large extractions, in-database execution, and analytical data structuring, which makes it easy for you to load big data into SAS Visual Analytics.

This paper examines how SAS Visual Data Builder is used and the benefits it brings to your self-service analytics and big-data environments.

DESIGN PRINCIPLES

SAS Visual Data Builder is a web-based interface to robust data querying and structuring. Simply put, it is a tool that generates Structured Query Language (SQL) code. Then, you can run that code on demand or schedule it to run at a later time. It is designed to give you the power of SQL queries, including the ability to execute inside your database relational system, without requiring you to know the SQL syntax. In addition, SAS Visual Data Builder has the built-in knowledge to load the output data into the SAS LASR Analytic Server.

The target audience for SAS Visual Data Builder in your organization is the business intelligence power user, the analytical modeler, and any user responsible for preparing data. It is not designed for activities typically performed by IT.

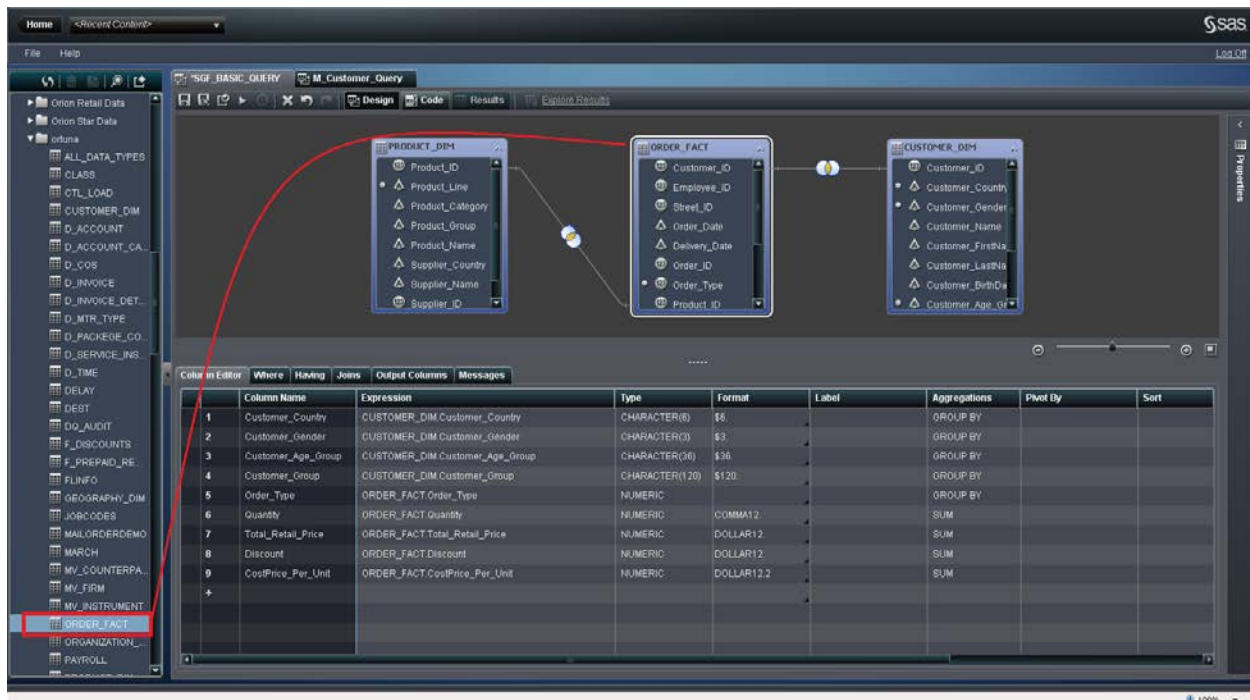
	SAS Visual Data Builder	Data Integration Tools
Capabilities	Simple tool with focused set of capabilities	Complex tool with wide range of capabilities
Permissions	Users need only limited set of permissions to the data	Need access to a larger set of enterprise data
Skills	Minimum contact with SQL	Programming and scripting skills are important
Sharing	Data created for a specific goal and shared with a small team	Data created for multiple goals and shared across multiple divisions

FEATURES

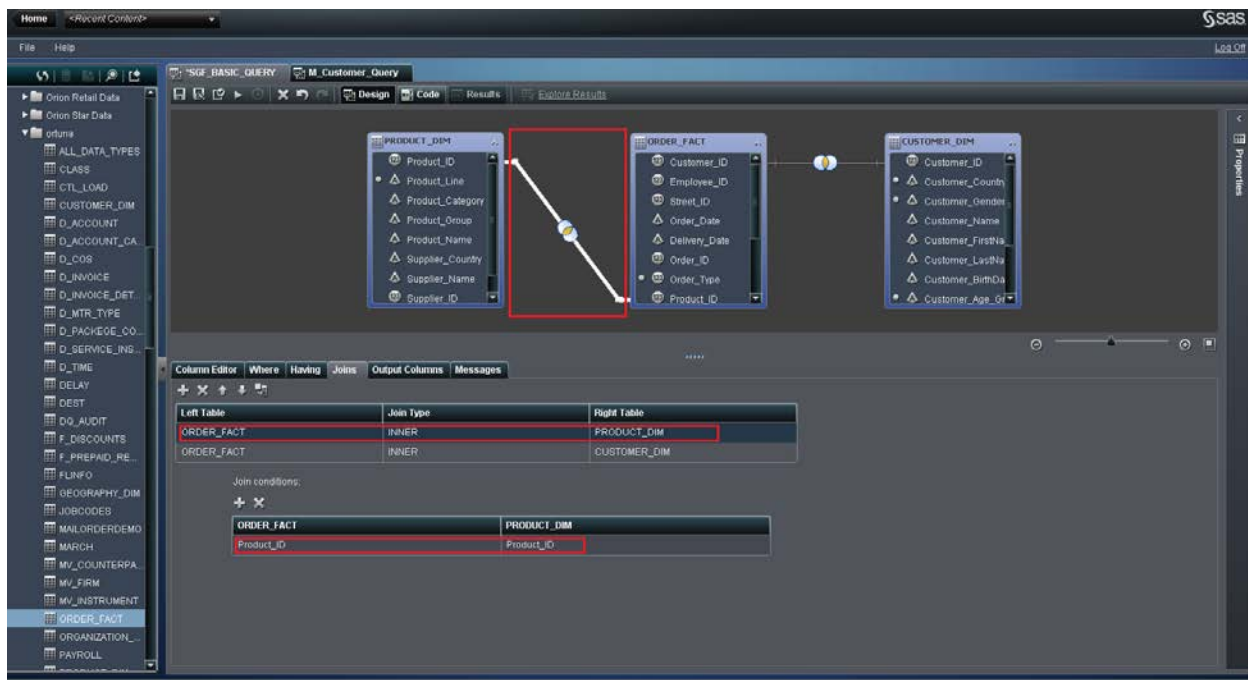
BASIC QUERY

SAS Visual Data Builder allows you to create an output table by transforming a single input table or joining multiple tables from the same library. Supported libraries and their tables must be registered in the SAS metadata before you can use them as query inputs. SAS Visual Data Builder allows you to save the entire structure of a query as a "Visual Data Query" and retrieve it at a later time.

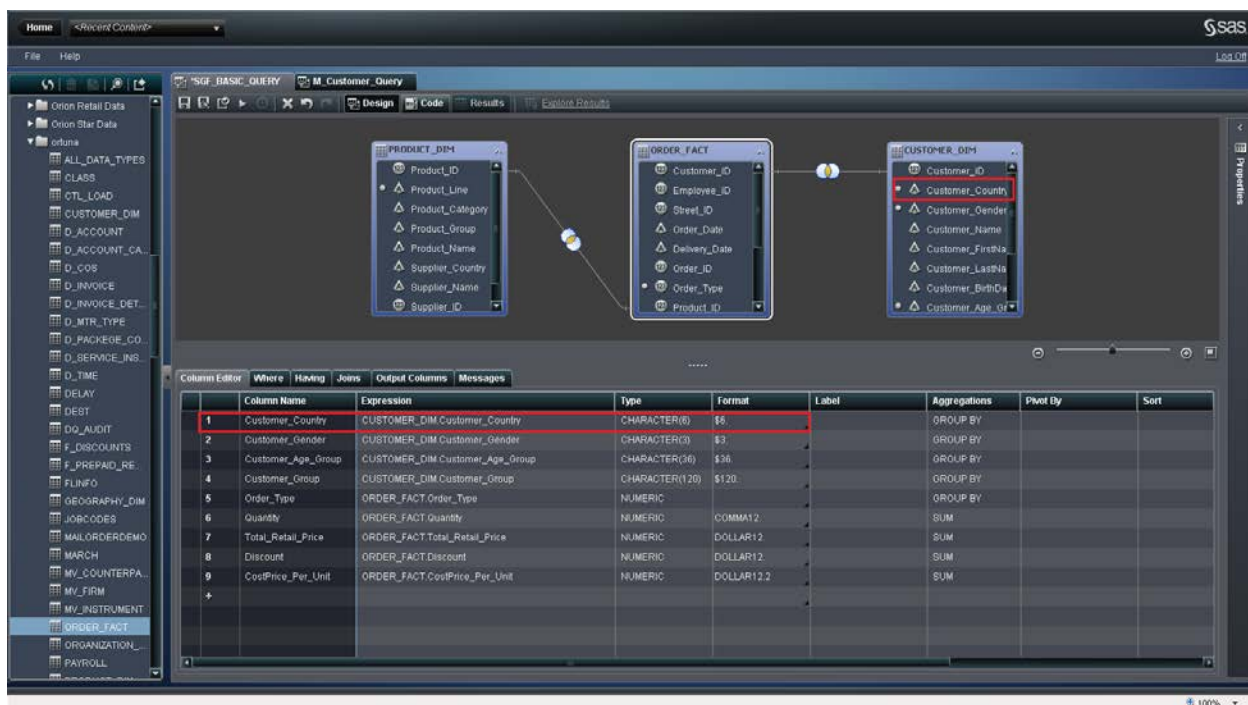
Displays 1 -9 walk through a typical query process. SAS Visual Data Builder allows users to drag and drop tables on to the palette, discover and modify joins, and manipulate the query output, including renaming columns, creating calculated columns with expressions, and filtering the data with complex logic. If landing the result to SAS LASR Analytic Server, you can explore the data with Visual Analytics Explorer with the click of a button.



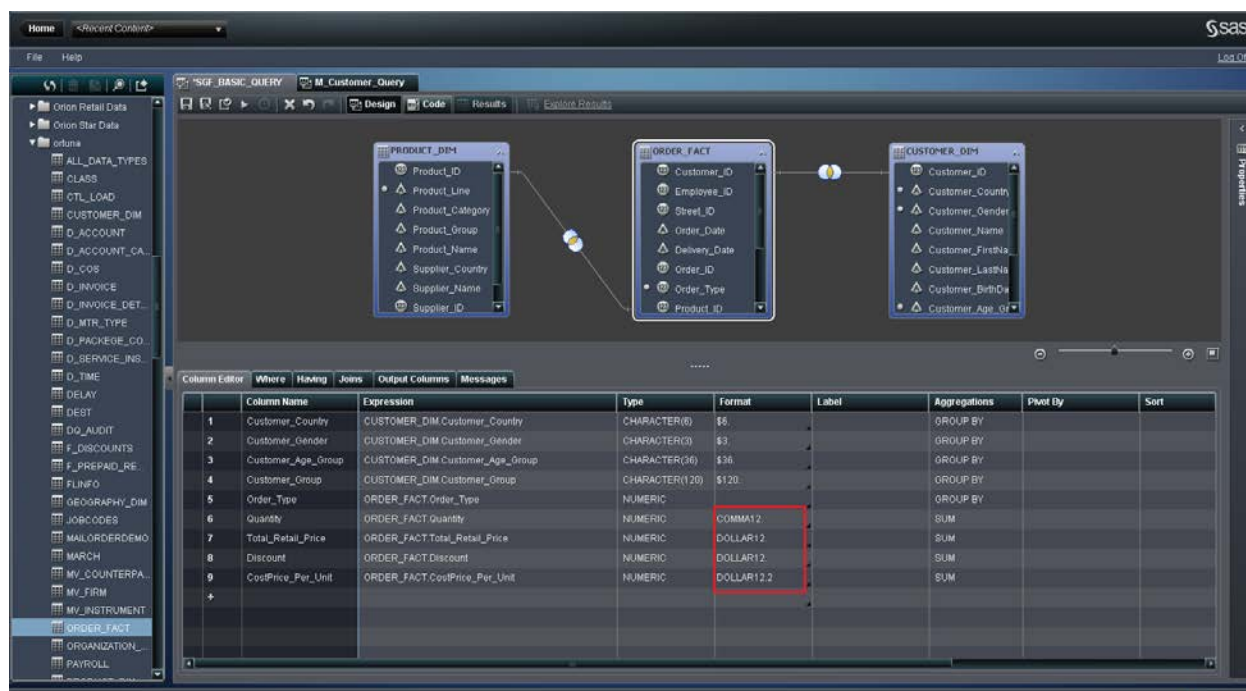
Display 1. To create a SAS Visual Data query, drag tables registered in the metadata.



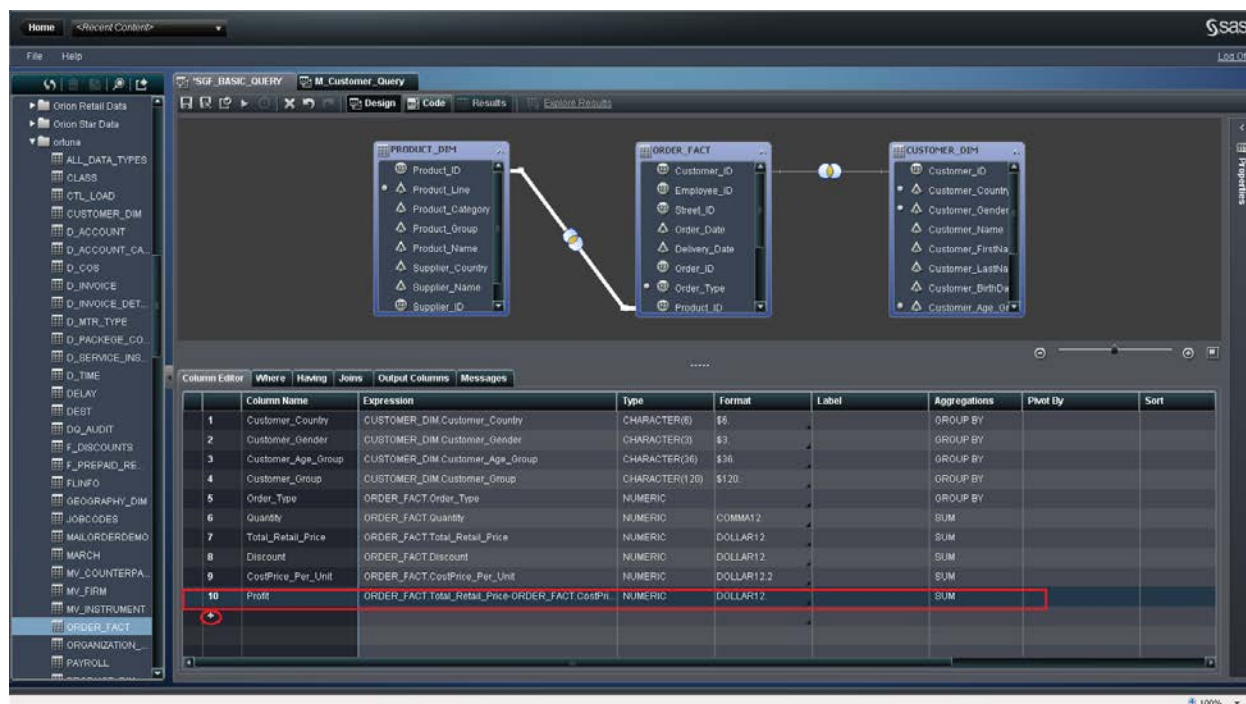
Display 2. Joins are auto-discovered, but you can modify them.



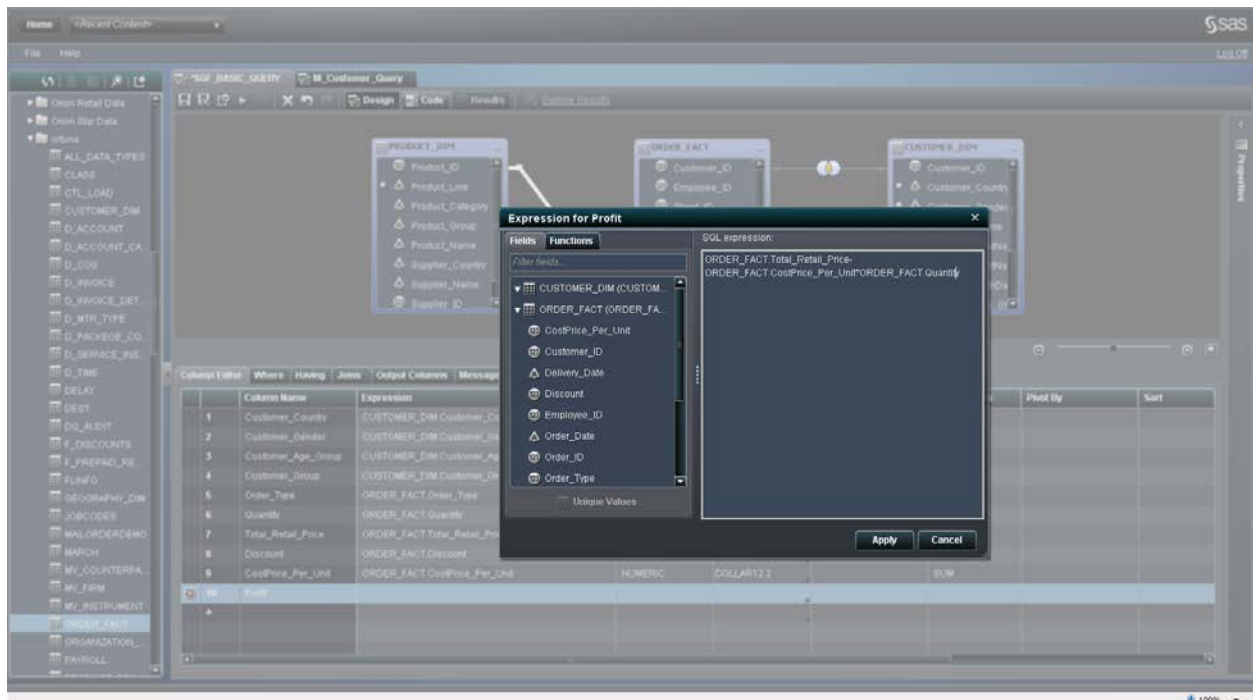
Display 3. To add columns to the query, click the columns inside the table.



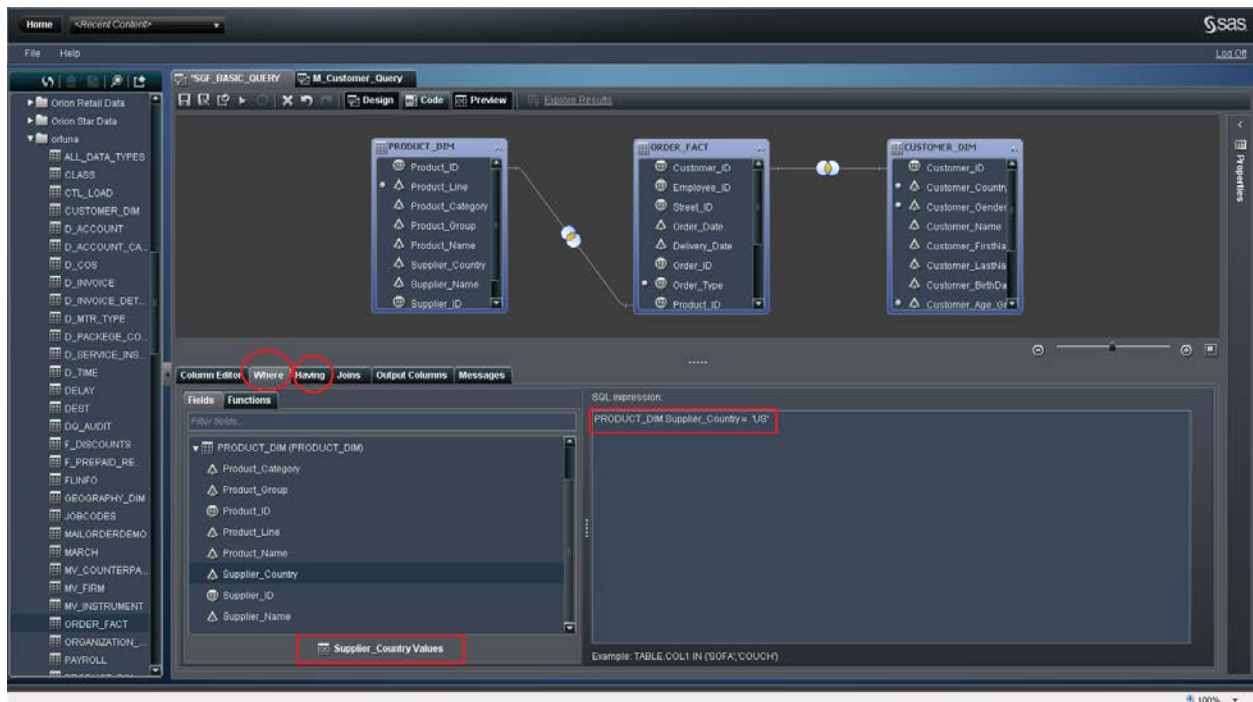
Display 4. To change column properties, such as formats, click on the format cell and select a standard SAS format.



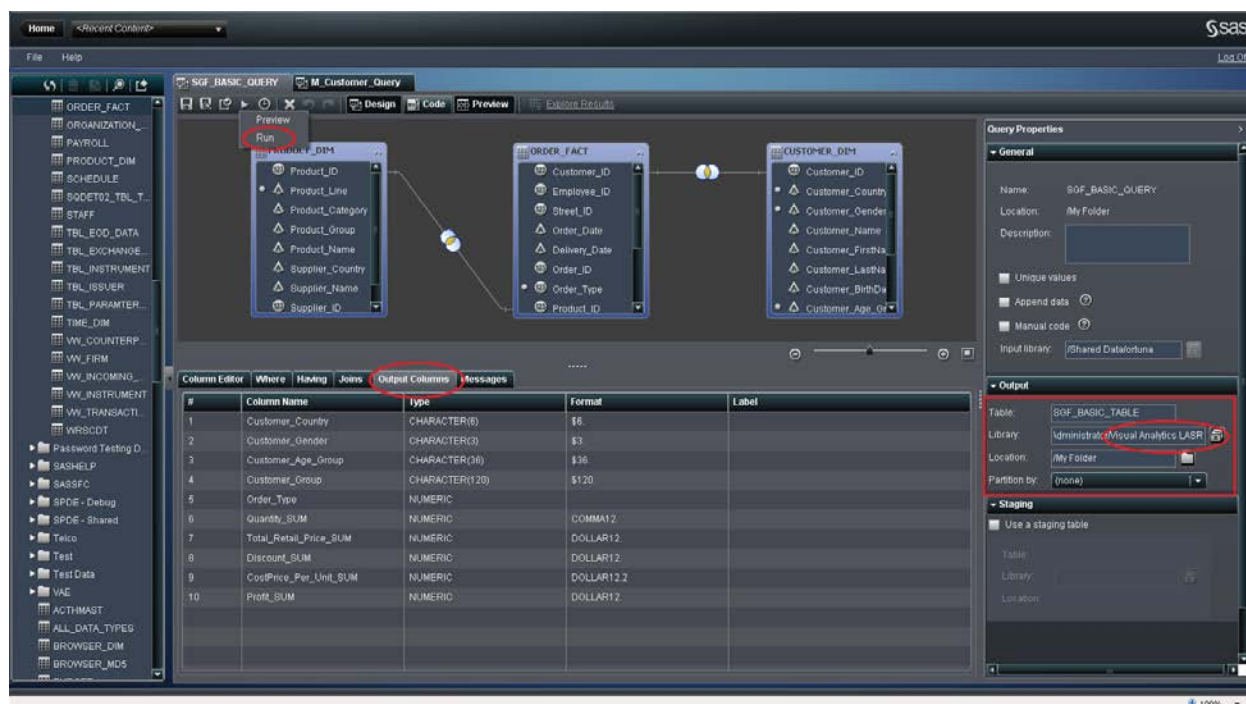
Display 5. To add a computed column, click inside the column editor.



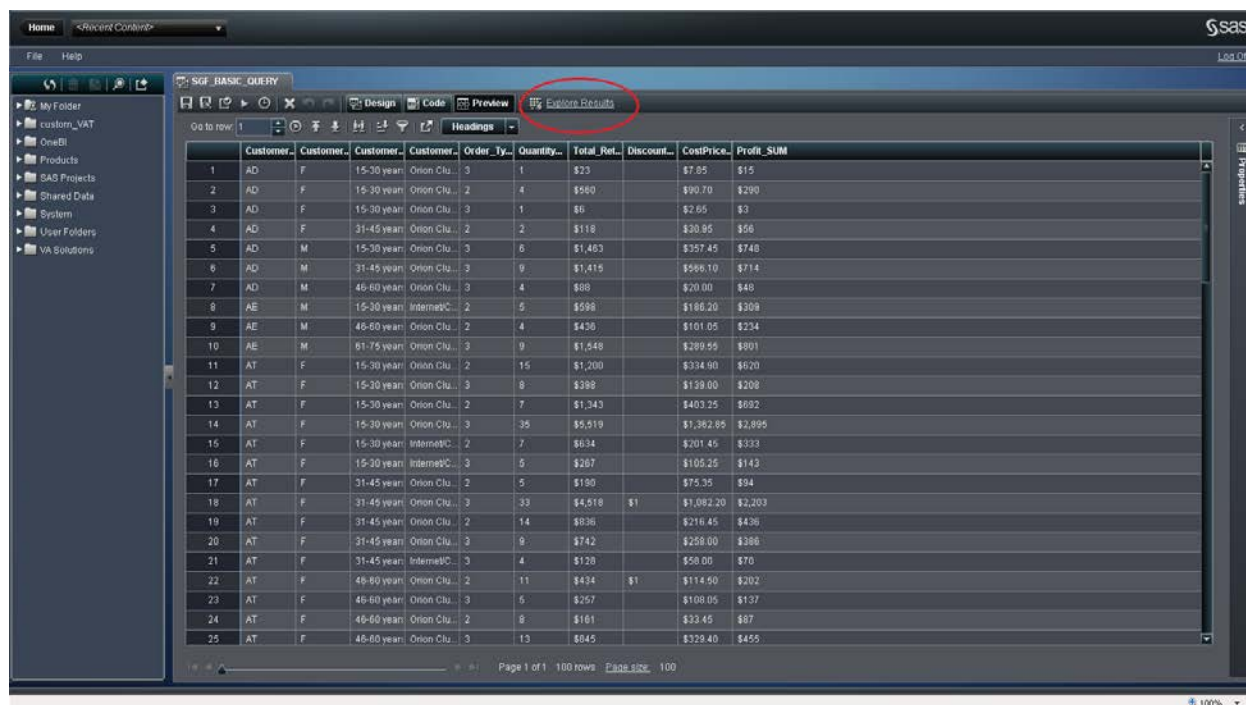
Display 6. To add a new expression for the column or modify an existing one, click the expression cell.



Display 7. To subset the data, add Where and Having filters.



Display 8. Add an output table and run the query. In this case, the output library is the SAS LASR Analytic Server.



Display 9. Preview the results. When the output is a LASR table, you can analyze the results in SAS® Visual Analytics Explorer.

SIMPLE SAS LASR ANALYTIC SERVER LOADING

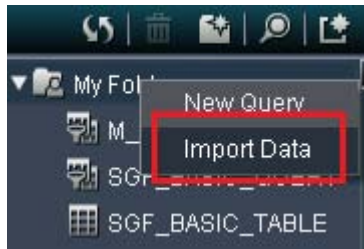
With a just a few clicks, SAS Visual Data Builder provides a streamlined way of loading four types of data:

- Comma-Separated Values (CSV) files
- Excel spreadsheets
- tables already defined in SAS metadata
- results of queries built in SAS Visual Data Builder

SAS Visual Data Builder registers the resulting table to metadata, and when the load is complete, SAS Visual Data Builder lets you quickly launch SAS Visual Analytics Explorer to begin immediately working with the data.

CSV Files

To load a CSV file, select the folder in which you want to store table metadata, right-click, and then select **Import Data**:



Display 10. Select **Import Data**.

Import Data

Source File

Name: CustomersWithIssues.csv

Delimiter:

☒ Comma ☐ Semicolon ☐ Space ☐ Tab ☐ Other

☒ Use data in the first row as SAS variable names

☐ Trim spaces

Data records begin on row: 2

Output Table

Name: CustomersWithIssues

Library: /Shared Data/LASR/AFIT/Libraries/BIRD English LASF

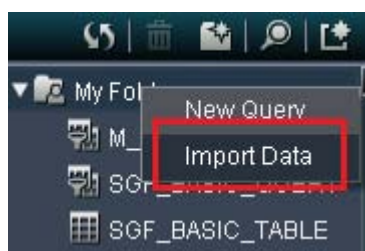
Location: /My Folder

Import **Close**

Display 11. Select a CSV file and type the output table information.

Excel Spreadsheets

The second option is to load an Excel spreadsheet.



Display 12. Select **Import Data**.

Import Data

Source File

Name: b1.xlsx

Select worksheet:

- Bal
- M
- Old

☒ Use data in the first row as SAS variable names

☐ Trim spaces

Data records begin on row: 2

Output Table

Name: BadIndicators

Library: /Shared Data/LASR/MAFIT/Libraries/BIRD English LASF

Location: /My Folder

Import **Close**

Display 13. Select an Excel spreadsheet and type the output table information.

Tables

Another option is to select a table that is already defined in SAS metadata and load it directly into memory.



Display 14. Select a table from the metadata tree, and select **Load a Table**.

Load a Table [X]

Specify the settings for loading the table to SAS LASR Analytic Server.

Source Table

Name:

Library:

LASR Table

Name:

Description:

Location:

Library:

Display 15. Type the output table information and submit.

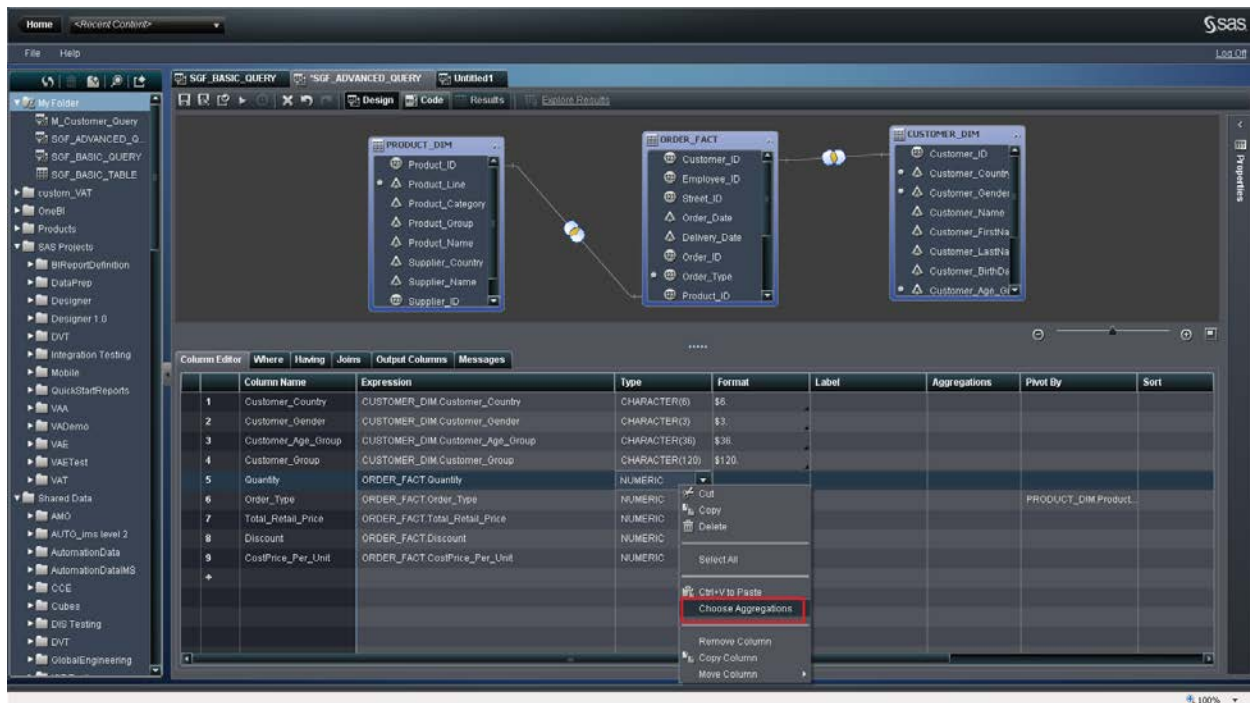
Query Results

Finally, users can build complex queries with SAS Visual Data Builder and land those results directly in SAS LASR Analytic Server. The basic query example above demonstrates that capability.

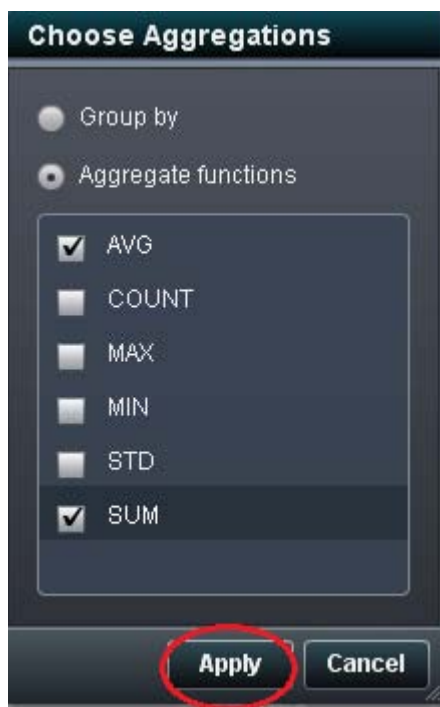
ADVANCED FEATURES

AGGREGATIONS AND PIVOTS

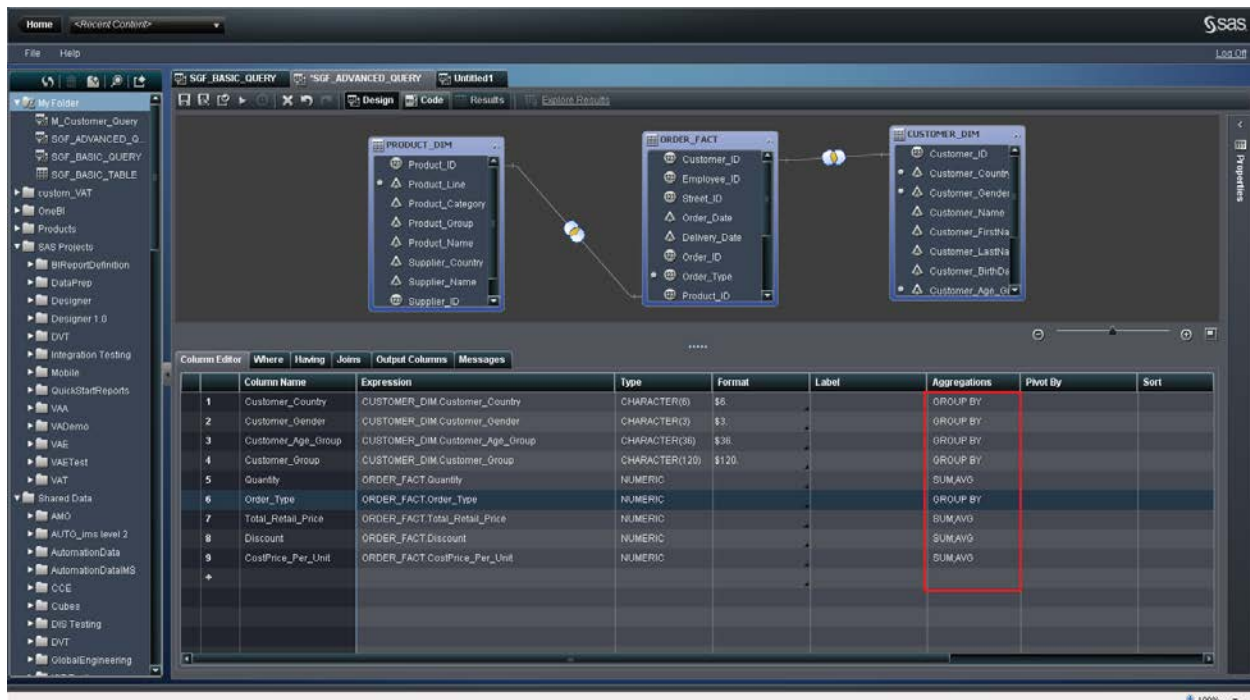
One of the most powerful features in SAS Visual Data Builder is the ability to aggregate data and pivot those aggregates across the values in another field.



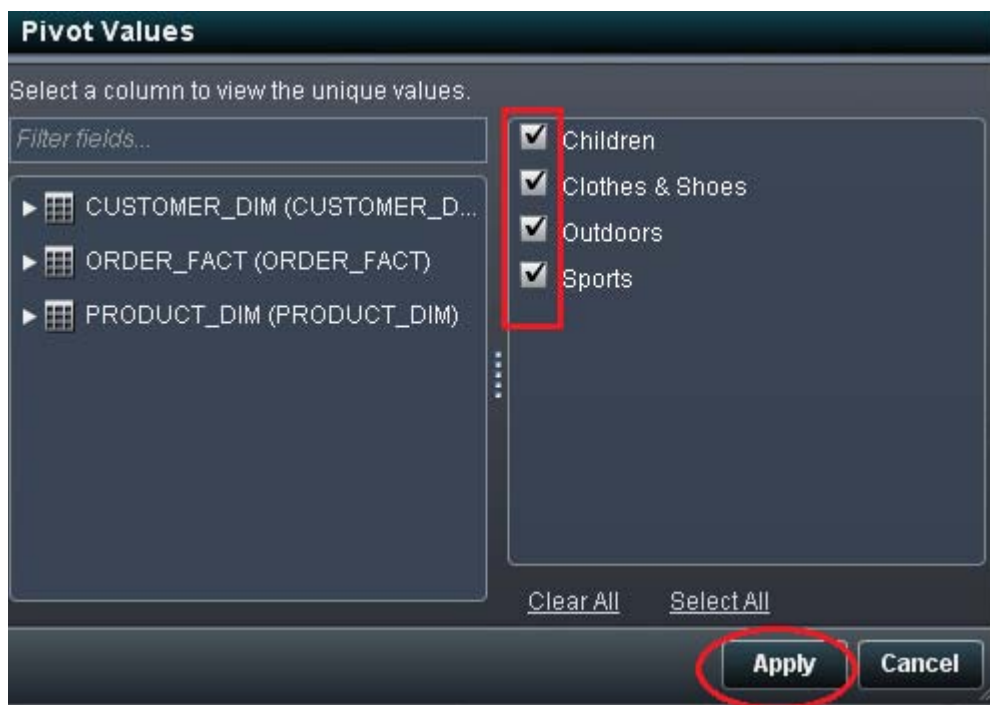
Display 16. Select one or more columns to select aggregation.



Display 17. Select the summary statistics for each aggregation, and click **Apply**.



Display 18. SAS Visual Data Builder automatically applies a group by to all the columns not selected for aggregation.



Display 19. Click the Pivot by for one of the aggregated columns. You can select a column and its values to pivot by.

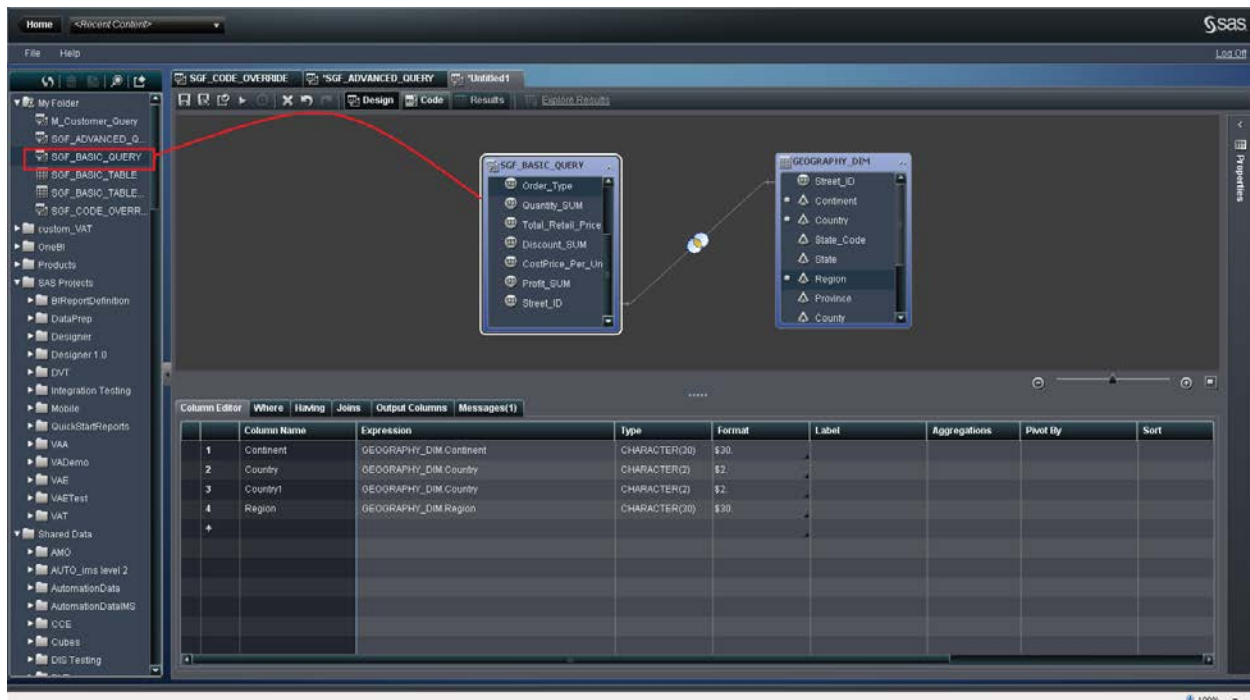
The screenshot displays the SAS Visual Data Builder interface. At the top, there's a navigation bar with 'Home' and '<Recent Content>'. Below it, a file explorer on the left shows a folder structure including 'My Folder', 'custom_VAT', 'OneBI', 'Products', 'SAS Projects', 'Shared Data', 'System', 'User Folders', and 'VA Solutions'. The main workspace shows a query design with three tables: **PRODUCT_DIM**, **ORDER_FACT**, and **CUSTOMER_DIM**. Joins are indicated between **PRODUCT_DIM** and **ORDER_FACT**, and between **ORDER_FACT** and **CUSTOMER_DIM**. Below the design view, the 'Column Editor' tab is active, showing a table of output columns. The table has columns for #, Column Name, Type, Format, and Label. Rows 7-14 are highlighted with a red box, showing aggregate columns for CHILDREN, CLOTHES__SHOES, OUTDOORS, and SPORTS categories.

#	Column Name	Type	Format	Label
1	Customer_Country	CHARACTER(6)	\$6.	
2	Customer_Gender	CHARACTER(3)	\$3.	
3	Customer_Age_Group	CHARACTER(36)	\$36.	
4	Customer_Group	CHARACTER(120)	\$120.	
5	Quantity_SUM	NUMERIC		
6	Order_Type_SUM	NUMERIC		
7	CHIL_Total_Retail_Price_AVG	NUMERIC		CHILDREN_Revenue_AVG
8	CLOT_Total_Retail_Price_AVG	NUMERIC		CLOTHES__SHOES_Revenue_AVG
9	OUTD_Total_Retail_Price_AVG	NUMERIC		OUTDOORS_Revenue_AVG
10	SPOR_Total_Retail_Price_AVG	NUMERIC		SPORTS_Revenue_AVG
11	CHIL_Total_Retail_Price_SUM	NUMERIC		CHILDREN_Revenue_SUM
12	CLOT_Total_Retail_Price_SUM	NUMERIC		CLOTHES__SHOES_Revenue_SUM
13	OUTD_Total_Retail_Price_SUM	NUMERIC		OUTDOORS_Revenue_SUM
14	SPOR_Total_Retail_Price_SUM	NUMERIC		SPORTS_Revenue_SUM
15	Discount_SUM	NUMERIC		
16	CostPrice_Per_Unit_SUM	NUMERIC		

Display 20. Notice the output columns created for each aggregate across pivot values.

SUBQUERIES

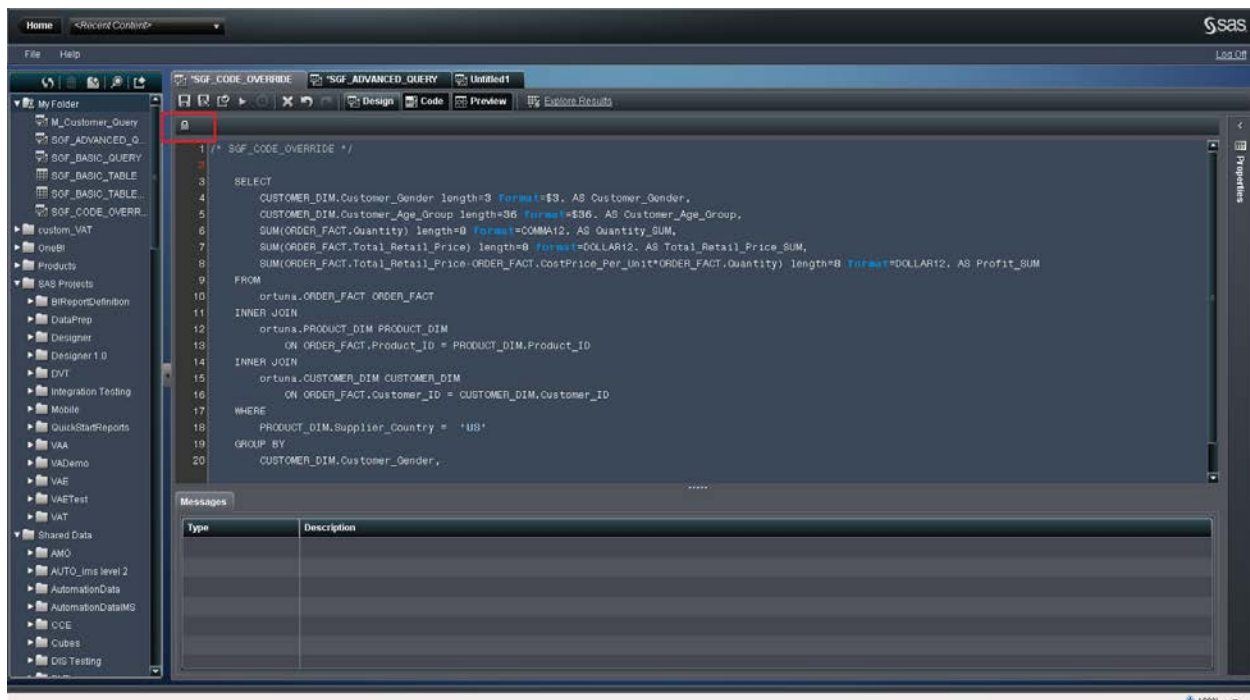
SAS Visual Data Builder allows for a query to be used within another query. This creates a powerful paradigm of nesting queries to satisfy many data structuring needs such as multiple fact-table joins.



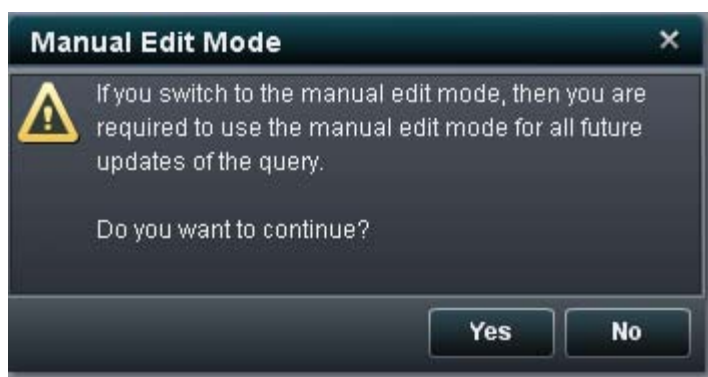
Display 21. Drag and drop another query to take advantage of this powerful feature.

CODE OVERRIDE

SAS Visual Data Builder allows you to view the SQL code it generates and, if needed, allows you to override that code.



Display 22. You can view the generated code and choose to override it by clicking the lock.



Display 23. Confirm the choice

CONCLUSION

In this paper, we have identified the key features of SAS Visual Data Builder and how they empower you to prepare and structure data from a wide range of sources without involving IT personnel.

ACKNOWLEDGMENTS

SAS Visual Data Builder is the result of collaboration between developers, testers, and product managers. The authors wish to thank everyone involved in this effort. This paper is dedicated to the memory of Lee Ballard who worked tirelessly to make sure this product is successful.

RECOMMENDED READING

- *SAS Visual Data Builder: User's Guide*
- *SAS Visual Analytics: Administrator's Guide*

CONTACT INFORMATION

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

Sam Atassi
SAS Institute Inc.
180 N. Stetson Ave, Suite 1600
Chicago, IL 60601
Work Phone: (312) 819-6800
E-mail: sam.atassi@sas.com

Malcolm Alexander
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Work Phone: (919) 531-2783
E-mail: Malcolm.Alexander@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.