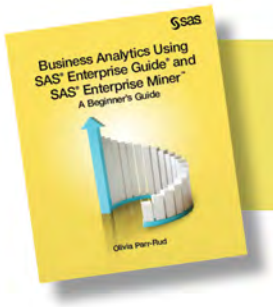
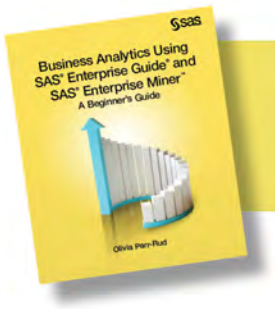
Business Analytics Using SAS® Enterprise Guide® and SAS® Enterprise Miner™

A Beginner's Guide

Olivia Parr-Rud



Contents



Chapter 6: Market Analysis Using SAS Enterprise Guide

Introduction

Chapters 1 through 5 focused on general knowledge and techniques that laid the foundation for many types of data analysis. In this chapter, you will explore a more specific topic: competitive analysis.

When your company sets its strategy, the first questions to answer are as follows:

- What are the strengths and weaknesses of our primary competitors as compared to us?
- What is our market share and how can we increase it?"

For most industries, data sources are available that allow companies to determine their market share, or "share of wallet," within certain segments of the population. This analysis is valuable for setting marketing strategies, guiding research and development, and informing finance and budget allocations.

Project Overview

The leadership team at DMR Publishing Company is interested in understanding the drivers of revenue within its business. It has gathered U.S. customer data for the past year that consists of revenues, numbers of publications, and three demographic variables. For this analysis, you will use SAS Enterprise Guide.

The project has eight steps:

1. Initiate the project in SAS Enterprise Guide 6.1.
2. Import and view market data.
3. Add the DMR Publishing customer SAS data set to the project.
4. Use the query tool to build a new age group variable.
5. Summarize the DMR Publishing customer data to a gender and age group level.
6. Merge customer and market data.
7. Summarize the merged data to age group level.
8. Perform penetration and a "share of wallet" analysis.

Market Analysis

So far in this analysis, our main interest has been to describe the characteristics of the customer base. If possible, it can be useful to compare characteristics of the customer database with the same characteristics in the overall market. This comparison enables a penetration analysis and share of wallet analysis as defined in Chapter 1.

We have purchased a data set from General List Company that shows the annual revenues and number of subscriptions for the entire publishing market within the United States, segmented by gender and age group. Our goal is to compare performance within gender and age group between the DMR Publishing's customer data and the total market data purchased from General List Company.

Project Initiation

Similar to the process in Chapter 5, the first step is to open the project. Double click the SAS Enterprise Guide icon. Select **New Project**.

Data Preparation

Your first step is to prepare the data.

Import Data

To access the data, go to **File ▶ Import** and choose the Publishing Market data set. When the first window opens, click **Finish**. Notice that the data is already summarized as shown in Output 6.1.

Output 6.1: Publishing Market Data

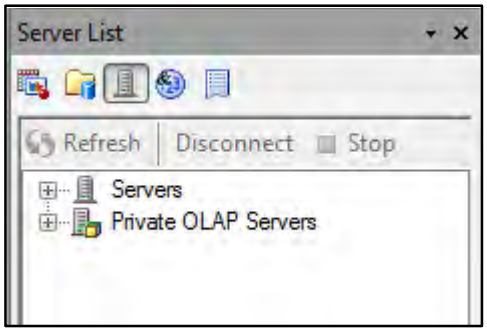
	SUBSCRIBERS	GENDER	AGE_GROUP	REVENUE	SUBSCRIPTIONS
1	10,484	F	18-21 Years Old	$431,319	24,187
2	14,169	F	22-25 Years Old	$885,233	43,451
3	11,753	F	26-30 Years Old	$904,011	37,418
4	16,077	F	31-40 Years Old	$1,303,178	51,428
5	11,421	F	41-50 Years Old	$1,103,420	40,624
6	9,432	F	51+ Years Old	$781,581	30,789
7	8,561	M	18-21 Years Old	$289,773	16,728
8	15,836	M	22-25 Years Old	$829,967	42,518
9	16,105	M	26-30 Years Old	$1,091,799	47,651
10	22,319	M	31-40 Years Old	$1,806,043	69,480
11	15,180	M	41-50 Years Old	$1,458,032	53,122
12	12,288	M	51+ Years Old	$1,156,802	42,818
13	1,243	U	18-21 Years Old	$45,998	2,609
14	1,566	U	22-25 Years Old	$87,298	4,398
15	1,174	U	26-30 Years Old	$76,911	3,411
16	1,149	U	31-40 Years Old	$83,765	3,497
17	696	U	41-50 Years Old	$59,602	2,258
18	571	U	51+ Years Old	$42,800	1,728

To create the penetration and wallet analyses, merge the market data to the DMR Publishing customer data. Before you can merge the two data sets, you need to summarize the customer data. We plan to merge the two data sets by gender and age group.

The GENDER field is ready to match, but you must create the AGE_GROUP variable in the DMR Publishing customer data.

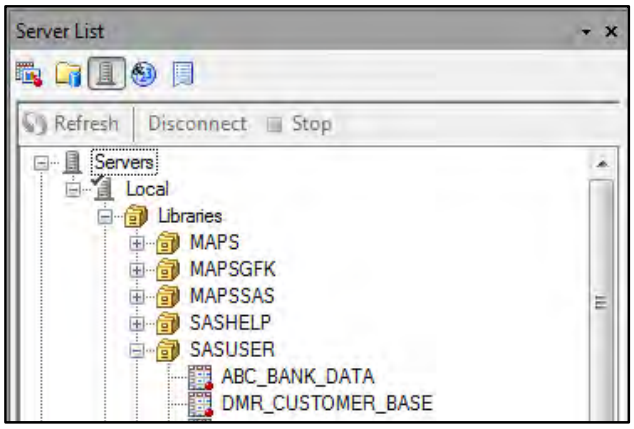
Because we started a new project, we need to bring the DMR Publishing customer data into the project. In the top menu, select **View ▶ Server List.** A new window will open below the Project Tree window; it will display the servers, as shown in Figure 6.1.

Figure 6.1: Access Server List to Locate DMR Publishing Customer Data



Click the plus sign to the left of **Servers.** Continue to expand until you reach the **SASUSER library** as shown in Figure 6.2.

Figure 6.2: Expand Library to Locate DMR Publishing Customer Data

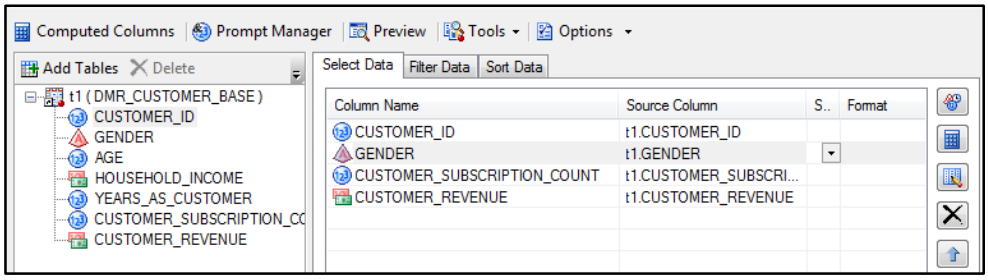


Right-click the DMR_CUSTOMER_BASE data and select **Open**. The data has now been added to the project and is visible in the Project Tree. You are ready to create the Age Group variable in the DMR Publishing customer data.

Build Age Group Variable

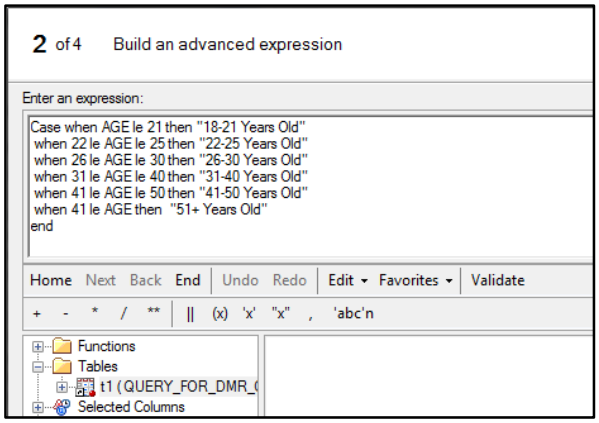
Right-click the DMR Customer data set and select **Query Builder** as shown in Figure 6.3. Drag four variables from the left over to the right column: GENDER, CUSTOMER_ID, CUSTOMER_SUBSCRIPTION_COUNT, and CUSTOMER_REVENUE.

Figure 6.3: Query to Build Age Group on DMR Publishing Customer Data



Next, click **Computed Columns ▶ New ▶ Advanced Expression ▶ Next** (Figure 6.4).

Figure 6.4: Expression Builder to Create Age Group Variable



The window has a space to enter an expression. Program 6.1 displays the syntax for building the variable in the Query Builder; copy it into the box under **Enter an expression.**

Program 6.1: Expression Builder to Create Age Group Variable

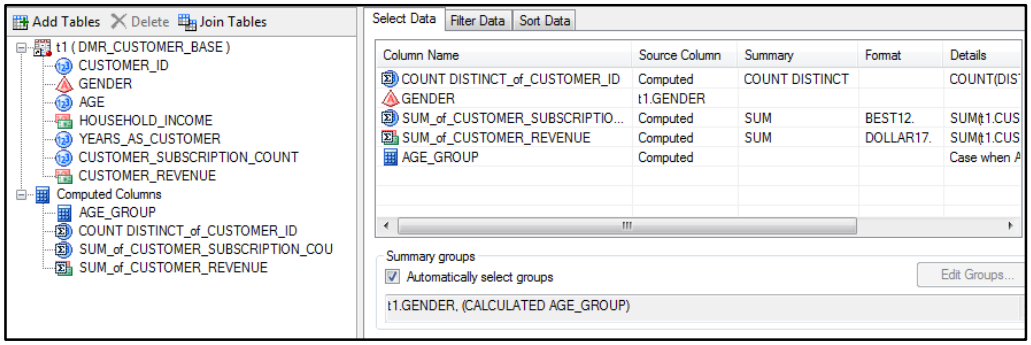
```
Case when AGE le 21 then "18-21 Years Old"
 when 22 le AGE le 25 then "22-25 Years Old"
 when 26 le AGE le 30 then "26-30 Years Old"
 when 31 le AGE le 40 then "31-40 Years Old"
 when 41 le AGE le 50 then "41-50 Years Old"
 when 51 le AGE then  "51+ Years Old"
end
```

Click **Next**. In the next window's field, type AGE_GROUP in the top box next to **Column Name**. Click **Finish ▶ Close.**

Sum Performance Variables

The next step is to sum the performance variables CUSTOMER_ID, CUSTOMER_REVENUE, and CUSTOMER_SUBSCRIPTION_COUNT (Figure 6.5). Within the same query, on the right-hand side, put your cursor next to CUSTOMER_ID under the **Summary** column. An arrow will appear. Scroll down and select COUNT DISTINCT. For easier viewing, you can adjust the column widths with your cursor. Next, right-click the space next to CUSTOMER_REVENUE and CUSTOMER_SUBSCRIPTION_COUNT and select SUM. You'll notice that a **Summary groups** pane appears at the bottom right section of the window. The software guesses GENDER and CALCULATED AGE_GROUP, which is correct. Click **Run**.

Figure 6.5: Sum Performance Variables across Gender and Age Group



In Output 6.2, the data set appears to be summarized correctly. The new summary variables have the prefix SUM_. We will use these variables in the penetration and wallet analyses.

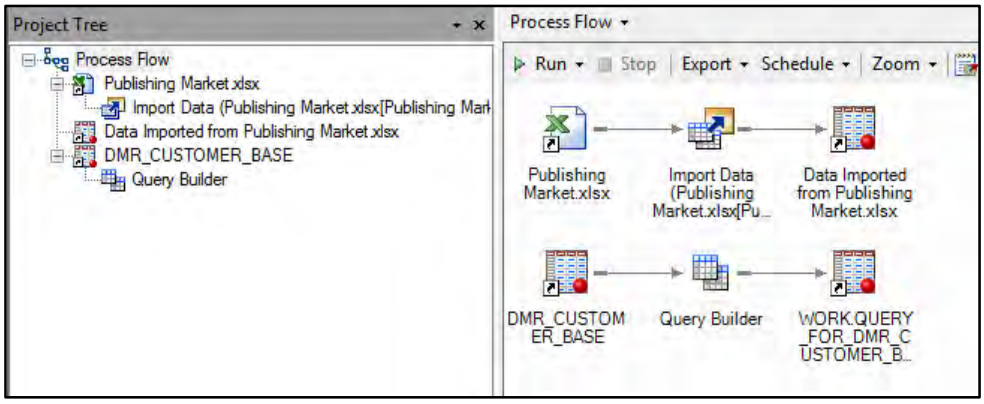
Output 6.2: Summarized DMR Publishing Customer Data

	COUNT DISTINCT_of_CUSTOMER_ID	GENDER	SUM_of_CUSTOMER_SUBSCRIPTION_COU	SUM_of_CUSTOMER_REVENUE	AGE_GROUP
1	606	F	3448	$64,845	18-21 Years Old
2	1036	F	4577	$80,295	22-25 Years Old
3	1119	F	4166	$78,140	26-30 Years Old
4	1550	F	5311	$105,465	31-40 Years Old
5	1138	F	4188	$80,985	41-50 Years Old
6	691	F	2240	$42,985	51+ Years Old
7	401	M	2012	$40,955	18-21 Years Old
8	1239	M	4944	$88,855	22-25 Years Old
9	1775	M	6514	$122,680	26-30 Years Old
10	2805	M	9853	$196,140	31-40 Years Old
11	1894	M	6970	$137,960	41-50 Years Old
12	1258	M	4198	$80,595	51+ Years Old
13	50	U	287	$5,310	18-21 Years Old
14	122	U	519	$9,315	22-25 Years Old
15	94	U	410	$8,480	26-30 Years Old
16	115	U	395	$8,105	31-40 Years Old
17	77	U	261	$4,910	41-50 Years Old
18	49	U	139	$2,620	51+ Years Old

Merge Customer and Market Data

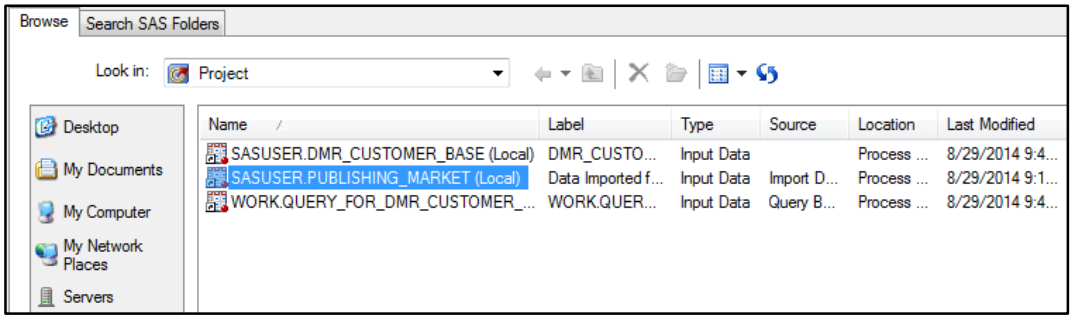
The next step is to merge the data. If you don't see the Process Flow on the right side of your screen, double-click on **Process Flow** in the top left corner at the top of the Project Tree, and it will appear in the work area on the right as seen in Figure 6.6.

Figure 6.6: Select Data for Merge Query



Right-click the newly summarized data and select **Query Builder**. After the window opens, drag all the variables from t1 (QUERY_FOR_DMR_CUSTOMER_BASE) into the right-hand column. In the upper left-hand corner, look for **Join tables**. A new window will open with the fields in the new data set. Click **Add Tables**. Select WORK.PUBLISHING_MARKET (Figure 6.7). If you can't see the whole name, put your cursor over the fields. The full name will appear. Click **Open**.

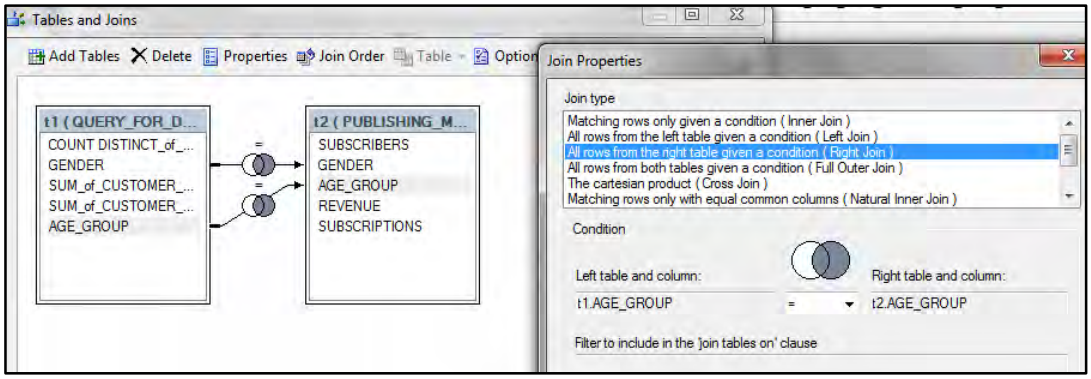
Figure 6.7: Select Data for Join



The software will guess a match variable. In this case, it guesses GENDER. But, it is an inner join as shown in the Venn diagram. Right-click on the circles and select **Properties**. You want to match the records in the DMR Publishing customer data that finds a match in the Publishing Market data; therefore, from the list at the top, select **All rows from the right table given the condition (Right Join).** Also, notice just below the circles that it asks if you want t2.GENDER to equal (=) t1.GENDER. The default is correct.

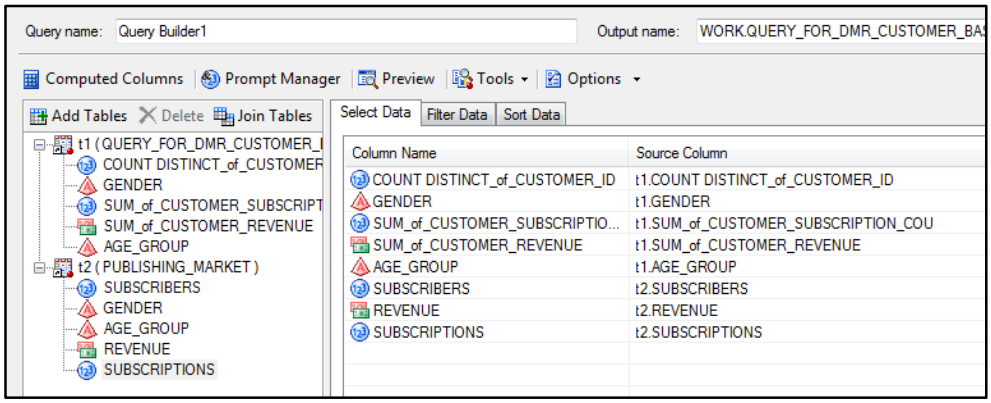
You are also going to match on AGE_GROUP. Right-click on AGE_GROUP in the first data set (t1). Follow the prompt to match to AGE_GROUP in the second data set (t2). Next, a window will appear, and a prompt will ask you what kind of join you want to select (Figure 6.8). As directed in the first join using GENDER, you want to match the records in the DMR Publishing customer data that finds a match in the Publishing Market data; therefore, select **All rows from the right table given the condition (Right Join)**. Also, notice just below the circles that asks if you want t2.AGE_GROUP to equal (=) t1.AGE. The default is correct. Click **Close**.

Figure 6.8: Merge of DMR Customer and Market Data Sets



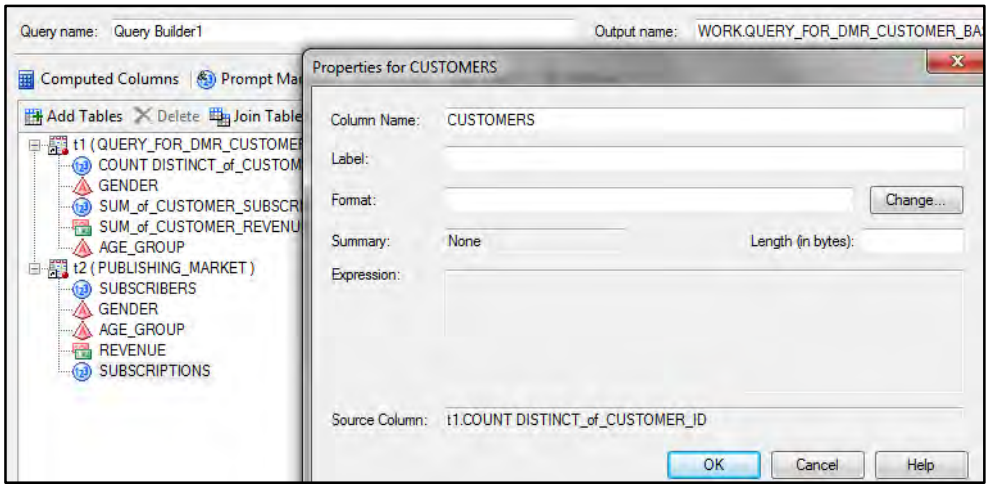
You will be back in the main screen of the query builder. From the left column, drag the following variables from the (t2) PUBLISHING_MARKET to the right side under Column Name: SUBSCRIBERS, REVENUE, SUBSCRIPTIONS (Figure 6.9).

Figure 6.9: Add Variables from Publishing Market Data



Before you run a query, you can change the column names (Figure 6.10).

Figure 6.10: Rename Three Columns to Business-Related Names



Double-click under **Column Name** on the right, where COUNT DISTINCT appears. When the window opens, type in a new name, such as CUSTOMERS. Click **OK**. Repeat the process for the two other summary columns. Suggested names are CUSTOMER_SUBS and CUSTOMER_REVENUE. Now, click **Run**.

The merged data shows a perfect match between the customer data and the market data. Notice that it is at the GENDER-by-AGE_GROUP level.

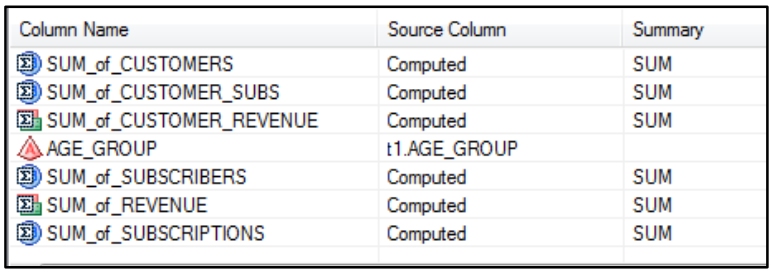
Output 6.3: Merged Data at Gender and Age Group Level

	CUSTOMERS	GENDER	CUSTOMER_SUBS	CUSTOMER-REVENUE	AGE_GROUP	SUBSCRIBERS	REVENUE	SUBSCRIPTIONS
1	606	F	3448	$64,845	18-21 Years Old	10,484	$431,319	24,187
2	1036	F	4577	$80,295	22-25 Years Old	14,169	$885,233	43,451
3	1119	F	4166	$78,140	26-30 Years Old	11,753	$904,011	37,418
4	1550	F	5311	$105,465	31-40 Years Old	16,077	$1,303,178	51,428
5	1138	F	4188	$80,985	41-50 Years Old	11,421	$1,103,420	40,624
6	691	F	2240	$42,985	51+ Years Old	9,432	$781,581	30,789
7	401	M	2012	$40,955	18-21 Years Old	8,561	$289,773	16,728
8	1239	M	4944	$88,855	22-25 Years Old	15,836	$829,967	42,518
9	1775	M	6514	$122,680	26-30 Years Old	16,105	$1,091,799	47,651
10	2805	M	9853	$196,140	31-40 Years Old	22,319	$1,806,043	69,480
11	1894	M	6970	$137,960	41-50 Years Old	15,180	$1,458,032	53,122
12	1258	M	4198	$80,595	51+ Years Old	12,288	$1,156,802	42,818
13	50	U	287	$5,310	18-21 Years Old	1,243	$45,998	2,609
14	122	U	519	$9,315	22-25 Years Old	1,566	$87,298	4,398
15	94	U	410	$8,480	26-30 Years Old	1,174	$76,911	3,411
16	115	U	395	$8,105	31-40 Years Old	1,149	$83,765	3,497
17	77	U	261	$4,910	41-50 Years Old	696	$59,602	2,258
18	49	U	139	$2,620	51+ Years Old	571	$42,800	1,728

For your current analysis, you want to look at penetration and share of wallet by age group. So, you will summarize the data one more time. Return to the **Process Flow** and right-click the merged data and select **Query Builder**.

When the window opens, drag each variable except GENDER to the right side and drop it under **Column Name**. Put your cursor in the empty space under **Summary**, and a dropdown menu will appear with the word NONE (Figure 6.11). Click and select SUM next to each variable except AGE_GROUP.

Figure 6.11: Summing of Merged Data to Age Group Level



Column Name	Source Column	Summary
SUM_of_CUSTOMERS	Computed	SUM
SUM_of_CUSTOMER_SUBS	Computed	SUM
SUM_of_CUSTOMER_REVENUE	Computed	SUM
AGE_GROUP	t1.AGE_GROUP	
SUM_of_SUBSCRIBERS	Computed	SUM
SUM_of_REVENUE	Computed	SUM
SUM_of_SUBSCRIPTIONS	Computed	SUM

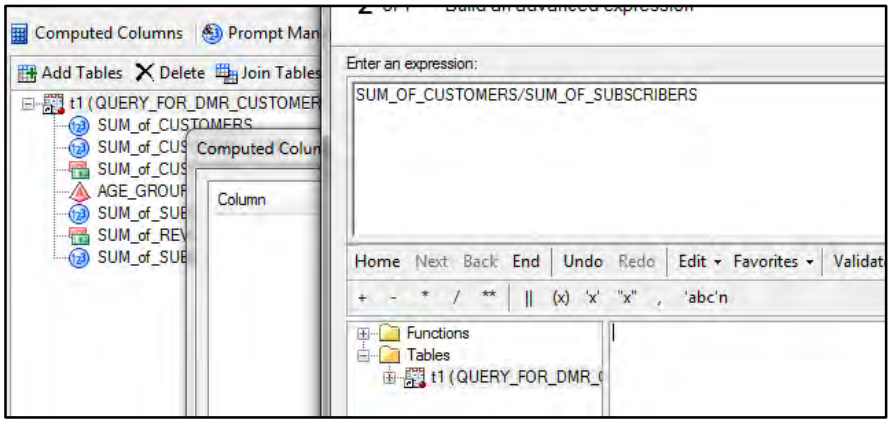
When you are finished, click **Run**. This step summarizes all the customer and market values to the AGE_GROUP level. Review and close the data view.

Calculate the Share of Wallet

The next step is to calculate the share of wallet for customers, subscriptions, and revenues with use of the query tool. With your cursor on the summarized data, right-click and then click **Query Builder**. Once it opens, pull all of the variables into the right-hand pane. You can pull each variable separately or move the entire data set at one time by dragging the icon where t1 (QUERY_FOR_DMR appears). For ease of viewing, widen the **Column Name** heading on the right side until you can see the full column names.

Next, go to **Computed Columns ▶ New ▶ Advanced Expression ▶ Next**. In the box under **Advanced Expression**, type SUM_of_CUSTOMERS/SUM_of_SUBSCRIBERS and click **Next (**Figure 6.12). Type PCT_of_SUBSCRIBERS next to **Column Name**. At the bottom of the window, next to **Format**, click **Change** and select **Numeric** and **Percentw.d**. Leave other settings unchanged as default. Click **OK ▶ Finish**. In the remaining window, click **Close**.

Figure 6.12: Computed Columns to Create Percentage Variables



Next, click **New** and type SUM_of_CUSTOMER_SUBS/SUM_of_SUBSCRIPTIONS. Click **Next** and name the fraction PCT_of_SUBSCRIPTIONS. Create the same format and click **Finish**.

Repeat the process with SUM_of_CUSTOMER_REVENUE/SUM_of_REVENUE in the box. Use the same format. Click **Next** and name the fraction PCT_of_REVENUE. Click **Next ▶ Finish**. Click **Close ▶ Run**.

> **NOTE:** To avoid errors, variable names must be exactly as stated.

Output 6.4 displays the final percentage values.

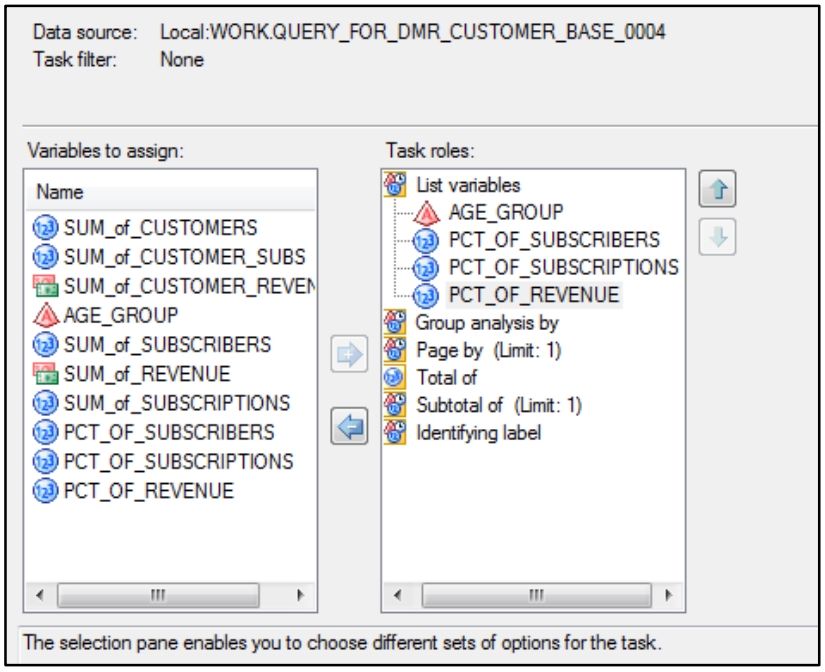
Output 6.4: Final Percentage Variables Showing Percent of DMR Subscriber Measures as Percent of Total Market

	PCT_OF_SUBSCRIBERS	PCT_OF_SUBSCRIPTIONS	PCT_OF_REVENUE
1	5%	28%	14%
2	8%	32%	10%
3	10%	38%	10%
4	11%	39%	10%
5	11%	42%	9%
6	9%	30%	6%

Penetration and Share of Wallet

Your data is now ready to analyze. To get a clear view by AGE_GROUP, close the data and highlight the last created data set. Click **Tasks ▶ Describe ▶ List Data.** This process allows you to take the variables or columns that you select and print them in a report format. The window in Figure 6.13 allows you to manage the report output.

Figure 6.13: Creation of a Report



Drag AGE_GROUP and the three percentage (PCT) variables to the right side. Click **Run**. Look for the output created in Microsoft Word.

Results

Reporting your results is one of the most important steps in your analysis. If you do everything right and can't communicate the results, your efforts and insights will be wasted.

In Output 6.5, the final report displays the market penetration analysis and share of wallet by age group. When you are evaluating the percentage of subscribers, the lowest market penetration is in the 18 to 21-year-old age group. It is higher with 26 to 50-year-olds, at 10% to 11%. It drops slightly at age 51 or older (9%).

Output 6.5: Final Penetration Analysis by Age Group

Row number	AGE_GROUP	PCT_OF_SUBSCRIBERS	PCT_OF_SUBSCRIPTIONS	PCT_OF_REVENUE
1	18-21 Years Old	5%	28%	14%
2	22-25 Years Old	8%	32%	10%
3	26-30 Years Old	10%	38%	10%
4	31-40 Years Old	11%	39%	10%
5	41-50 Years Old	11%	42%	9%
6	51+ Years Old	9%	30%	6%

The percentage of subscriptions shows a different outcome. The youth, aged 18 to 21 years, have a higher number of subscriptions per person. So, they bring in the largest percentage of the revenue. Their share of wallet is 14%, compared with only 6% for those aged 51 or older.

Several conclusions can be drawn from this analysis. If you are looking for additional subscribers that bring in high average revenue, the 18- to 21-year-old market is the place to focus. These subscribers seem to be happiest with the current offerings, although there is still much room to grow.

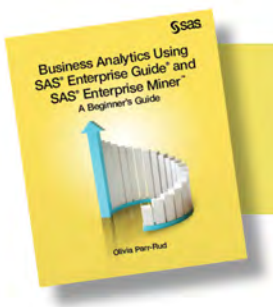
On the other hand, growth opportunities for the business may emerge in the other age groups if new products are emphasized. The share of wallet is 10% or less for customers aged 22 years or older. This result suggests that products could be developed to appeal to groups older than 21.

Notes from the Field

This case study provides some rich insights and potential opportunities for DMR Publishing. What the company does with the information depends much more on the abilities of the organization.

Consider that there are two new opportunities identified. The first opportunity is to find additional subscribers that look like their current subscriber base. The job of finding additional subscribers might be handled by the marketing department. The second opportunity is to create new products for the underserved areas of the market. This need for new products would likely be a project for the product development side of the business. The optimal solution may be a mixture of both departments.

Because both market growth for the current business and product development are important, management must make the decision about how to allocate resources. The best outcomes are seen when companies have fostered a culture of big data so that departments are able to collaborate and partner in the implementation of the analysis results, optimizing profits for the overall company.



Index

A

B

C

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U

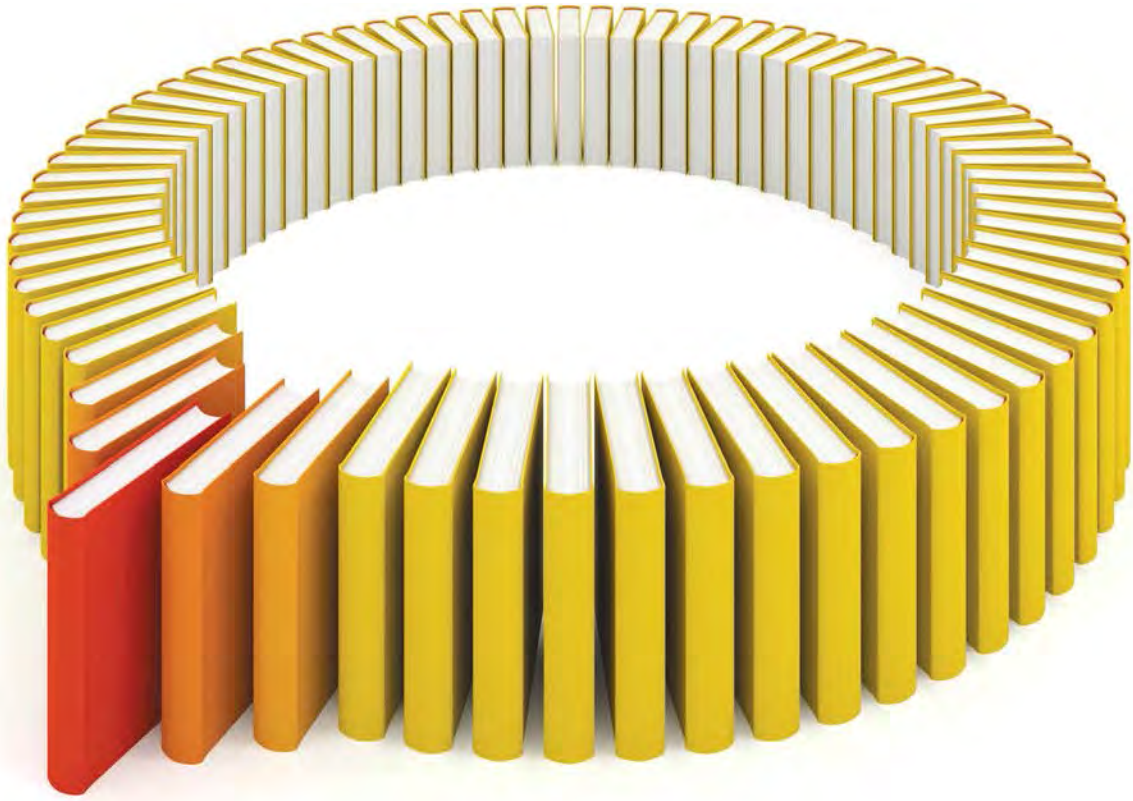
V

W

About the Author

Olivia Parr-Rud, an internationally recognized expert in predictive analytics, business intelligence, and innovative leadership, founded the SAS Data Mining Users Group, having been a SAS user since 1991 and a SAS instructor and conference presenter for many years. She hosts the popular VoiceAmerica Business radio show *Quantum Business Insights* (http://www.voiceamerica.com/show/2240/quantum-business-insights) and is a thought leader in the integration of analytic tools and leadership practices to optimize performance and organizational agility. Her predictive analytics research founded her first book, *Data Mining Cookbook: Modeling for Marketing, Risk and Customer Relationship Management* (Wiley 2000), which unveiled links between the global economy and organizational dynamics, a topic featured in her second book, *Business Intelligence Success Factors: Tools for Aligning Your Business in a Global Economy* (Wiley and SAS Institute Inc. 2009). Her current research will develop a model that tracks the alignment of business intelligence with human intelligence. Parr-Rud holds a B.A. in mathematics and an M.S. in statistics. In addition to public speaking, she offers training and consulting in both predictive analytics and business leadership, with clients that include numerous Fortune 500 companies. For more, visit http://oliviagroup.com.

Learn more about this author by visiting her author page at http://support.sas.com/publishing/authors/parr-rud.html. There you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more.



Gain Greater Insight into Your SAS® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

support.sas.com/bookstore
for additional books and resources.

sas
THE POWER TO KNOW®