Business Intelligence with SAS Enterprise Guide 4

Program

1. About SAS EG
2. Opening SAS EG
3. How to make and store a dataset
4. How to open and import
5. Filter and Query
6. Formats
7. Basic descriptive statistics
8. Graphs
9. Regression analysis
10. OLAP cubes
11. Break
12. Assignment 1
13. Assignment 2
14. Assignment 3

About SAS EG	SAS Enterprise Guide is a graphical user interface (GUI) building upon SAS BASE. SAS EG can be viewed as a tool for business intelligence and incorporates many of the same features as Microsoft Excel, SPSS etc.	
Opening SAS EG	SAS EG is located in *All programs → SAS → Enterprise Guide 4*.	The user interface looks like most other GUI programs for Windows. Some of the most important elements of the interface for you to be aware of in this respect is the 'Project Designer', 'Project Explorer' and 'Cube View Manager'.
How to make and store a data set	In SAS EG new data sets are created by following the path *File → New → Data*.	In SAS EG libraries are referring to particular folders on the hard disk.

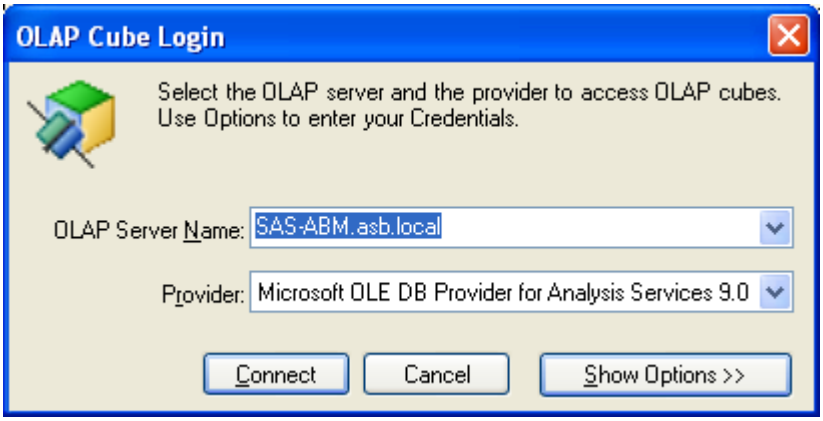
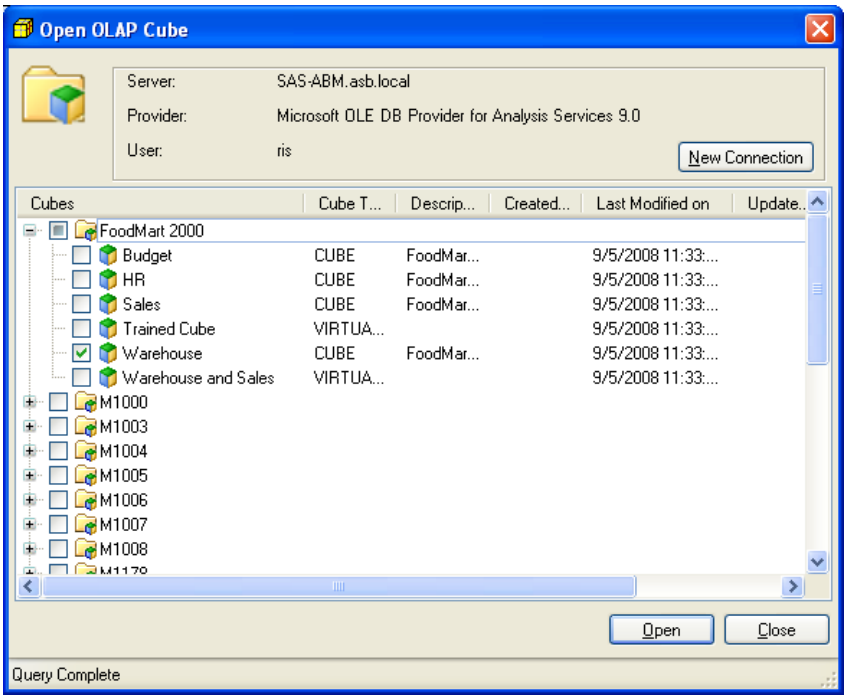
	The only peculiar characteristic to be aware of is in which library the data set is stored. By default it is stored in the temporary library called *Work*, which is cleared when the program is closed. Therefore it is important to ensure that you store the new data set in another location at the end of the session.	It is possible to create new libraries following the path *Tools → Assign Library*.
How to open/ import and export data sets	Existing data sets can easily be imported no matter if it is a SAS data set or a data set from Excel. It is done by *File → Open → Data* or *File → Import → Data* Here you will have to choose whether your data is stored on the local computer or on a SAS server. If you wish to export your data sets (or tables/graphs from different analyses) to other formats like .xlsx for Excel, it can be done through *File→Export→Export Data*	When using the import function it is possible to change what to include in the data set viewed in the project designer. You can e.g. exclude variables or observations from the original data set.
Filter and Query (Data manipulation)	With *Filter and Query* it is possible to filter, sort or compute variables. Every time a *Query* is made, a new data set is created. The functions are accessed through *Project Designer → Right click the data set →Filter and Query*.	Filtering and sorting data is done by *Filter Data* and *Sort Data* respectively. If you wish to compute a new variable, it is done through the *Computed Columns* function.
Basic descriptive statistics	To get an overview of a new data set it is often a good starting point to have a look on some basic descriptive statistics. In SAS EG this is accessed through *Describe → Summary Statistics* In this menu you will have a lot of options regarding which variables to get a description of and which measures to include	

	in the output.	
Graphs	Like the functions available for generating descriptive statistics the graphing functions are quite intuitively build. The different possibilities can be accessed through *Graph* → ... It is recommended to make use of the Graph Wizards as they ensure you get through all relevant settings for the specific graph.	
Regression analysis	Regression analysis is an typical example of how to perform a statistical analysis in SAS EG. The predefined analyses are grouped in *Analyze* →,.. where *Regression* is one of the possibilities. In the *Regression* window you will have to choose the dependent variable along with explanatory variables. Furthermore it is possible to check the necessary conditions like normally distributed and uncorrelated errors, no perfect multicollinearity and variance homogeneity through the functions accessible under the headings *Statistics* and *Plots* in the left part of the Regression window.	As Enterprise Guide is building upon SAS BASE it is possible to make your own customized analyses through the coding tool. Remember all the functions in Enterprise Guide are just a graphical representation of functions coded in SAS BASE
OLAP cubes OLAP terminology in SAS EG	*OLAP cube* An OLAP cube is essentially a multidimensional table of data. Furthermore the dimensions of the cube can be hierarchical ordered.	As an example you can have sales data for a supermarket chain with the dimensions *store sales, costs, geography, product categories* just to mention some of the possibilities.

	Dimensions Dimensions are logical categories in which the data gets grouped.	Typical examples include changing whether the sales level should be shown for the product dimension or the store dimension or contrary whether you want to see sales or stock levels for the different stores.
	Slice and dice A popular expression covering the possibility to quickly switch between different views of the aggregated data, e.g. to change between dimensions and measures.	E.g. to see sales level for the whole chain → classified by country → classified by county → classified by city → classified by shop
	Drill down Drill down covers the possibility to see a given measure sorted by different levels of aggregation.	
Navigating OLAP cubes	Navigating OLAP cubes in SAS EG is quite straightforward in the default view in the OLAP Viewer.	
(Drill down and drill up)	The '+'-mark is used to drill down in the shown dimension	
Changing content of the Cube View (Slice)	The downward pointing arrow just beside '+' is used to slice the data in the case you want to focus on only some of the data provided in the cube.	
(Dice)	To change the dimensions shown you can navigate the Cube View Manager for the desired dimensions and/or measures and then add them to the row or coloumn of the current	

	view. The Cube View Manager is by default shown in the left part of the screen and can be shown/hidden by the 'View Manager'-button in the Cube View menu bar.	
Removing dimensions or measures from the current view	Dimensions and measures can be removed from the current view by right clicking the dimension/measure and choosing "remove ..... from table"	
Creating measures	If you want to calculate some new measures from the existing data, it can be done by *Measure → Add measure → ...* where you can choose between different more or less advanced calculations.	As an example you can create a contribution margin measure from existing turnover and cost information by choosing *Measure → Add measure →Simple calculations → Percent Increase*
Creating filters	Creating filters can be used when you want to filter your data based on a boundary for the measure shown in the current view. An example could be to filter away all stores or supermarkets with sales less than 10.000 $ per week.	The difference between slicing and creating filters is essentially that slicing give you a complete slice of your cube of data where applying filters pick out the relevant observations from not only one slice of the cube but instead from the whole cube.
Bookmarks	When working with huge amounts of data in an OLAP cube you often have to modify the cube view a couple of times before you reach the view of relevance to the actual Business Intelligence case. To avoid carrying out all these steps every time you shall use the same cube view, you can bookmark the current view. This is done by	

	using the *Bookmark → Add bookmark →* … function in the Cube Viewer Menu Bar	
Highlighting measures	As a user of OLAP cubes you will probably often search for either unusually high or unusually low measures. To avoid having to look through every single measure in a cube view, you can highlight the measures above or below a specific boundary. This is done using *Highlight → Add highlight → …* path in the Cube Viewer Menu Bar.	

Accessing the OLAP cube for the assignments	The 2nd of the below assignments is supposed to be solved using the Warehouse cube accessed through *File → Open → OLAP Cube →*  *Connect →*  *→ Open*
Assignment 1 – Exploring SAS EG thorugh a traditional data set	Import the file bibliotekspenge999.xls. → bib 1. Filter out all the writers who earns less than 100,000 2. Make a new variable named tax (which is 40% of income). 3. After making the tax variable, make a new variable named deduction, which is calculated with the following formula: deduction = $tax^2/(income + 50.000)$ 4. Subtract deduction from tax and calculate. Name

	the variable Sum 5. Sort the data by Name 6. Save the manipulated data set on your personal desktop.
Assignment 2 – OLAP cubes	Open the Warehouse-cube as described above. By default the dimensions of the cube should be time, invoiced amount and products, shown by "Year". "StoreInvoice" and "All Products" on the aggregate level. As a manager of FoodMart 2000 you are interested in how well the different American stores in the chain are performing, therefore you need to navigate the data summarized in the OLAP cube to find the relevant views. 1) Manipulate the view of the cube so you see the Warehouse Sales and Profit for years 1997 and 1998 on the aggregate level for all shops in the USA, Regarding the product dimension you have to drill down to the level of "Product Departments". (Hint: Use Data Dimensions, Filter and Drill down) 2) Note down the aggregate profit and sales for Breakfast Foods in USA for both 1997 and 1998. What is the profit margin in the two years for this Product Department? 3) If you didn't choose to do so in 2), calculate a new measure termed "Profit Margin" as "Warehouse Profit" / "Warehouse Sales" and add the new measure to the cube view. 4) Slice the cube view to show only data regarding 1998. 5) Using "Conditional highlights", highlight and note down the 5 most profitable Product Departments in the USA as a whole. 6) Drill down to the state of California (CA) and apply the same highlight you just created in 5) with the slight difference that you only want the top 3 departments now. Which 3 Product Departments are the most profitable in this state? 7) As a final step, drill up to the USA as store level and down to quarters as time measure and isolate the

	third quarter of 1998. Bookmark this view.
Assignment 3 – Descriptive statistics and regression analysis (Optinal)	Open the file nissan.sas7bdat 1. Make descriptive statistics of the variable "km" and include a histogram 2. Make a scatter plot with km and price 3. Calculate a new variable named age, that measures the age of the Nissan 4. Make a regression with price as dependent variable including all the other variables