

# Stat 657

## Assignment 06 - SAS

### Scope:

This assignment will primarily utilize, but is not limited to, techniques covered in lectures 5 through 9 but does not require the use of any of the SQL set operators. You will practice using joins, subqueries, inline views and summary functions. You will not be creating any data sets in this assignment. Librefs to Orion and homework data must be protected with readonly access. Use a filename statement to define the path to the PDF output file.

### Specific Instructions for this Assignment:

1. Create a report entitled “2003 NCAA Team Scoring Analysis”, from the **scholarship03** dataset that you downloaded for last week’s assignment. **This is the same report as the report in step #4 from assignment 05 with a new twist**—you will get the overall average PPG from an inline view instead of a table created in another step. Everything you need is in **scholarship03**.
  - a. Inside the inline view you will need to compute the overall average PPG of players in the tournament. Remove those whose team had a seed of 15 or 16. This inline view only returns the company average salary.
  - b. There must be only one row for each team included in the report. This must be done WITHOUT remerging summary statistics back in with the original data and WITHOUT using the DISTINCT (or UNIQUE) keyword.
  - c. In addition to the team name, the report shows the number of players on the team roster in this data set, the average PPG (points per game) of the players on the team, and a column that represents the team average PPG as a percentage of the overall average (avg\_ppg/avg\_PPG\_all).
  - d. Create a final column, labeled PPG Level, with a value of “Above Avg.” when the team average PPG is greater than the overall average. Otherwise, the column has a value of “Avg. or Below”.
  - e. Since some teams in the data set did not include a complete roster, we only want to include teams that had 5 or more players in this report.
  - f. Arrange the output by average team PPG from highest to lowest.
  - g. It is normal to get a note in the log about a Cartesian product.
2. Your co-worker has been maintaining a list of givers and their donations in an Excel file. Since you may be using a SAS environment that is not configured to import Excel, the file has been converted for you to a dataset called **givers**. This data set is in the Assignment Data Files folder on eCampus and should be saved to the permanent library you are using to store homework data. NOTE: Disregard any notes in the log about Cross Environment Data Access. This is a result of using a Linux system or 64 bit Windows to access data created on an older PC system. In addition to data sets in your homework library, the remaining steps in this assignment will

also use Orion data sets provided with the SQL course notes you downloaded previously from eCampus.

Your boss suspects that the **givers** file is a mess and has asked you to provide some reports to assess the situation and help clean up the data. These reports will be described below.

3. Create a list of records from **givers** with duplicate names as shown in the example. HINT: Use a summary function and a Having clause. (It is OK to get the REMERGE message on this step.)
4. Create a list of Active Employees who are not in the giver list (based on employee\_id). Names are found in orion.employee\_addresses. The employee\_term\_date can be read from the orion.employee\_payroll table. Use a subquery in the where clause to determine which IDs to eliminate.
5. Use data in one or more of the tables above to create a list of people from the **givers** table who are no longer active employees at Orion Star. Show the ID, Name, and Gender of terminated employees.
6. Create a report entitled “Orion’s Customers Who Bought Products Other Than Shoes” using a multiway join. Include customer\_id, customer\_name, customer\_address, country, product\_group, and Birth Month. Get the information you need from orion.order\_fact, orion.customer, and orion.product\_dim. List customers who purchased products from product\_groups that are not designated as being shoes (Know thy data and pay attention to detail on this part). Sort the list by country, birth month, customer name, and product group.
7. Three PDF files must be uploaded to WebAssign. The program must contain a completed header block and comment blocks for each step. Options must be set so the output looks like that in the Assignment output posted on eCampus.