SAS Training: Company Transaction File

The purpose of this assignment is to build on concepts learned in the "Student Grades" assignment (e.g. RETAIN, FIRST-dot, LAST-dot, OUTPUT, etc.) to construct a company HR transaction file from various data sources.

For this assignment, you will need to read-in three data files. Since these files will be set together, it is recommended that you use INFILE so that the variable lengths will be consistent across the files. For your convenience, these files are already in CSV format. They are: "Munchies – HR Transactions.csv", "Munchies – Pay Scales.csv" and "Munchies – Location Data.csv". The first file contains the "HR" transactions for about a dozen employees. Included in this file are "HIRE" records, "TERM" records, and "POS" records, which indicate promotions and job title changes. This file also contains the position held by each employee, as well as EFFDATE, which represents the date each listed transaction occurred. The second file contains the pay scales for each position within the Munchies Empire. The pay rate for each position changes roughly every year. Finally, the third file contains the store location of each employee. Most of the employees are located at Munchies' main location (#1), but occasionally employees will change to a different restaurant.

Of course, Munchies has been sued, so it is time to start analyzing the data. The first thing that the client wants is for you to build a single master file that contains all of the relevant information from the three files. To begin…

1. First, read the three data files into SAS using INFILE.

2. Next, the client wants to you put each employee's store location on to the main HR file. Since store changes are not recorded in the HR file, you will need to stack the HR file with the locations file. This may mean that extra records will be added. After running the appropriate sort, you should use RETAIN to "drag" down the store location on to the records that do not already have a store location included. In the end, you should have a clean file that has only one new variable (store location) along with a few new records for employees who changed store locations. For these new records, you should recode the missing RECTYPE variable to equal "LOC". **It is important to note that when using RETAIN that you must not "cross into" the next employee's records. You should always have a BY statement and reset the RETAIN variable using "FIRST-DOT" logic.**

3. Next, you need to bring the pay rates on to each of the records in the HR file. The client insists that you put the correct pay rate on the file, exact to the *day*. It is not sufficient to merge the pay records on by YEAR or even by YEAR and MONTH. (Hint: to get started, you should stack the pay file and the HR file and then sort by POSITION and EFFDATE and then make use of RETAIN to "drag" the pay rate down). Make sure that

the file is cleaned-up once you've completed this step, such that you have only added 1 new variable—the day/position-specific pay rate for each employee.

4. Next, using one sort and one data step (NO merging for now!), you should use RETAIN to create a new variable called "HIREDATE". The hire date for each employee should be consistent across all records.

5. Next, using one sort and one data step, create a variable called "TERMDATE". The termination date for each employee should be consistent across all records. If the employee is still active, TERMDATE should be equal to 12/31/2999. Both TERMDATE and HIREDDATE should be formatted using the mmddyy10. format. At this point, your data file should look like the original HR data set with 4 new variables, LOCATION, PAYRATE, HIREDATE, and TERMDATE. You may also have some added records due to employee location changes.

6. The class period for the case is from 2001 to 2008. The attorney wants to get an employee head count on monthly basis for each month during the class period. In order to accomplish this, the attorney asks that you create a "month-end incumbent file". This involves running through the transaction file you have created with a DO loop and using "OUTPUT" to create a record for each month an employee is active.

7. Finally, using the month-end incumbent file that you have built, create a monthly summary file that includes head counts for each job title. The file should have 5 variables: MONTH, SERVERS, KITCHEN, LOCATIONMGR, ASSTMGR, where each of the last 4 variables has the respective head counts for each position. It should have 8 years' worth of records (8 x 12 = 96 records total).