DA 6223 Exam I

Due March 1, 2020

Instruction:

Please follow the instructions below to complete this exam.

- The SAS Enterprise Guide project (.egp) file needs to be turned in **through Blackboard before 11:59 PM CT on the due day**. **No late** exam will be graded.
- The **SAS Enterprise Guide project (.egp) file needs to be formatted as Firstname_Lastname_ABC123_EXAM1.egp (For example, if a student is named Wenbo Wu, his SAS Enterprise Guide project file should be named Wenbo_Wu_ABC123_EXAM1.egp)**. **Failing to name the project file in the required way will result a 5% deduction on the grade.**
- You should **create separate Process Flows** for each problem. **Failing to do so will result a 10% deduction on the grade.**
- **For Problem 5**, please use the Note function in SAS Enterprise Guide to write down answers for each part.
- **All solutions must be written up independently, without looking at another student's work.**

Business Problem: A national veteran's organization seeks to better target its solicitations for donation. By soliciting only the most likely donors, less money is spent on solicitation efforts and more money is available for charitable concerns. Solicitations involve sending a small gift to an individual and include a request for a donation. Promotions to donors include mailing labels and greeting cards.

The organization has more than 3.5 million individuals in its mailing database. These individuals are classified by their response behaviors to previous solicitation efforts. Of particular interest is the class of individuals identified as *lapsing donors*. These individuals made their most recent donations between 12 and 24 months ago. The organization seeks to rank its lapsing donors based on their responses to a greeting card mailing sent in June of 1997. (The charity calls this the 97NK Campaign.) With this ranking, a decision can be made to either solicit or ignore a lapsing individual in the June 1998 campaign.

The source of this data is the Association for Computing Machinery's (ACM) 1998 KDD-Cup competition. The data set and other details of the competition are publicly available at the UCI KDD Archive at `http://kdd.ics.uci.edu/`.

Problem 1 (10 points): Define SAS Library
For each student, the homework data files are saved under

O:\MSDA2020\DA6223_002\Instructor\Data\Exam 1\StudentID

where the StudentID is formatted as X_Y, for X to be the first name of the student, and Y to be the ABC123 ID number. For example, the StudentID for Wenbo Wu, will be Wenbo_ABC123. Define a working library **hw1** for above directory and **limit the access to be readonly**.

Problem 2 (20 points): Import Demographics Data
The demographics_StudentID.txt data is a fixed width text file that contains the demographical information of the donors. The names and descriptions of the columns of this data are as follows:

Table 1: Variables in Demographics Data

Name	Description
ID	Control Number
DemAge	Age
DemCluster	Demographic Cluster
DemGender	Gender
DemHomeOwner	Home Owner
DemMedHomeValue	Median Home Value Region
DemMedIncome	Median Income Region
DemPctVeterans	Percent Veterans Region

Use the Import Data task in SAS Enterprise Guide to complete the following:

(a) Change the output SAS dataset name to **demographics** and save this file in the WORK library.

(b) Change the variable names in the output SAS data set using the names in Table 1 and label them with the descriptions in Table 1.

(c) Change the length of DemGender and DemHomeOwner to be 1.

(d) Change the type of ID to be character.

(e) Do NOT embed the data and rename the task as **Import Demographics**.

(f) Use the **Summary Statistics** task to obtain the following statistics: mean, standard deviation, minimum, maximum, and number of missing values for DemAge, DemMedHomeValue, and DemMedIncome by different levels of DemGender. Based on the SAS Enterprise Guide result, summarize your findings into a table.

Problem 3 (20 points): Import Gifts Data
The gifts_StudentID.csv data is a common delimited file that contains the gifts the donors have made in the past. The names and descriptions of the columns of this data are as follows:

Use the Import Data task in SAS Enterprise Guide to complete the following:

Table 2: Variables in Gifts Data

Name	Description
ID	Control Number
GiftAvg36	Gift Amount Average 36 Months
GiftAvgAll	Gift Amount Average All Months
GiftAvgCard36	Gift Amount Average Card 36 Months
GiftAvgLast	Gift Amount Last
GiftCnt36	Gift Count 36 Months
GiftCntAll	Gift Count All Months
GiftCntCard36	Gift Count Card 36 Months
GiftCntCardAll	Gift Count Card All Months
GiftTimeFirst	Time Since First Gift
GiftTimeLast	Time Since Last Gift

(a) Change the output SAS dataset name to **gifts** and save this file in the WORK library.

(b) Label the variables with the descriptions in Table 2.

(c) Change the type of ID to be character.

(d) Do NOT embed the data and rename the task as **Import Gifts**.

(e) Use the **Summary Statistics** task to obtain the following statistics: mean, standard deviation, minimum, maximum, and number of missing values for GiftAvg36 GiftAvgAll, GiftAvgCard36, GiftAvgLast, GiftCnt36, GiftCntAll, GiftCntCard36, and GiftCntCardAll. Based on the SAS Enterprise Guide result, summarize your findings into a table.

Problem 4 (13 points): Import Promotion Data

The promotion SAS data contains the promotions have given to donors in the past and the current donation status. (**Hint:** To import a SAS dataset into the project, please use Query Builder and include all columns) The names and descriptions of the columns of this data are as follows:

Table 3: Variables in Promotion Data

Name	Description
ID	Control Number
PromCnt12	Promotion Count 12 Months
PromCnt36	Promotion Count 36 Months
PromCntAll	Promotion Count All Months
PromCntCard12	Promotion Count Card 12 Months
PromCntCard36	Promotion Count Card 36 Months
PromCntCardAll	Promotion Count Card All Months
StatusCat96NK	Status Category 96NK
StatusCatStarAll	Status Category Star All Months
Target_B	Target Gift Flag
Target_D	Target Gift Amount

(a) Change the output SAS dataset name to **promotion** and save this file in the WORK library.

(b) Use the **Summary Statistics** task to obtain the following statistics: mean, standard deviation, minimum, maximum, and number of missing values for PromCnt12, PromCnt36, PromCntAll, PromCntCard12, PromCntCard36, PromCntCardAll, and Target_D. Based on the SAS Enterprise Guide result, summarize your findings into a table.

(c) Use the **One-Way Frequency** task to obtain the frequency counts for each level of StatusCat96NK, StatusCatStarAll, Target_B. Based on the SAS Enterprise Guide result, summarize your findings into a table.

Problem 5 (57 points): Filter and Sort, Query Builder
In order to answer the following questions, you need to use Filter and Sort, Query Builder, or any other related tasks to get information from the three imported data sets.

(a) What is the ID numbers for the youngest male and female donor(s) in the data set **with known age and gender**?

(b) How many males and females have made donation this time for donors **with known gender**? (Hint: count the total number of 1's for Target_B.)

(c) What is the average age for male donors and female donors who made at least 3 donations in the past 36 months (**GiftCnt36**) for donors **with known gender**?

(d) Create a SAS data set in WORK library named "Improved_Donor" which includes all the **demographics**, **gifts** and **promotion** information for donors whose *Gift Amount Average 36 Months* is higher than *Gift Amount Average All Months*. How many observations do you have in the resulting data set?

(e) List the ID numbers for the Donor who has been the donor for the longest time and how long (in terms of number of months) has him/her been a donor?

(f) Who (list the ID) made the most recent donations? If multiple made the most recent donation at the same time, **only list the ID** for who made the largest most recent donation amount. What is the most recent donation amount for that donor?

(g) Create a SAS data set in WORK library named "Target_Donor" which includes the donors whose *Gift Amount Average All Months* is higher than or equal to $20, and *Promotion Count All Months* is less than 50, and who made the recent donation for at least $20. How many observations do you have in the resulting data set?

Problem 6 (15 points): Split Columns

(a) Create a SAS data set in WORK library named "Donation_Gender_Status96NK", including DemGender, StatusCat96NK, and average Target_D for males and females only (not for unknown gender donors).

(b) Use the Split Columns task to generate table such that the columns are DemGender and average Target_D for each 96NK status category.

(c) Use "Average_Donation_" as the prefix for the new columns.

(d) Save the output data in the WORK library and call it "Average_Donation_by_Gender".

Problem 7 (15 points): Stack Columns

(a) Create a SAS data set in WORK library named "Promotion_Gender", including DemGender, average PromCnt12, average PromCnt36, average PromCntAll, average PromCntCard12, average PromCntCard36, and average PromCntCardAll for males and females only (not for unknown gender donors).

(b) Use the Stack Columns task to generate table such that the columns are DemGender, CountType, and Count.

(c) Save the output data in the WORK library and call it "Promotion_Gender_Stack".