Due Date: See Canvas for the deadline

Reminder: All homework solutions must be completed independently. The SAS Enterprise Guide project (.egp) file must be turned in **on Canvas before the due date**. The **SAS Enterprise Guide project (.egp) file needs to be formatted as Lastname_Firstname_ABC123_HW1.egp.** You should also **create separate Process Flows** for each problem.

Business Problem: A national veteran's organization seeks to better target its solicitations for donations. 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 including a donation request. 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 solicit or ignore a lapsing individual in the June 1998 campaign.
The source of this data is the Association for Computing Machinery (ACM). The data set and other details of the competition are publicly available at the UCI KDD Archive at http://kdd.ics.uci.edu/.

Use the SAS EG or PROC SQL, or both to complete the tasks. As long as they're suitable for each task, choose the tool(s) you feel comfortable with.

Problem 1 (5 points): Define SAS Library

Download the data files shared with you and save them in a folder named Homework1, i.e., under P:\Homework1. *Failing to save the files exactly in this path will result in -5 points.*

Define a working library **hw1** for the directory you store the files in and **limit the access to be read-only**.

Problem 2 (15 points): Import Demographics Data

The **demographics.txt** data is a fixed-width text file containing the donors' demographical information. The names and descriptions of the columns of this data are given in Table 1.

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
DemMedIncome	Median Income

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. (2 pts)
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. (2 pts)
c) Change the length of DemGender and DemHomeOwner to be 1. (2 pts)
d) Change the type of ID to be character. (2 pts)
e) Do NOT embed the data and rename the task as **Import Demographics**. (2 pts)

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.

Bb Q1) Based on the SAS Enterprise Guide result, summarize your findings into a table: Right-click on the HTML output of the Summary Statistics and select **Share > Export**. Save the resulting HTML output in your P:\drive to be uploaded on Canvas. (5 pts)

Problem 3 (10 points): Import Gifts Data

The **gifts.csv** data is a comma-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:

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

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

a) Change the output SAS dataset name to **gifts** and save this file in the WORK library. (1 pts)
b) Label the variables with the descriptions in Table 2. (2 pts)
c) Change the type of ID to be character. (2 pts)
d) Do NOT embed the data and rename the task as **Import Gifts**. (1 pts)
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.

Bb Q2) Based on the SAS Enterprise Guide result, summarize your findings into a table: Right-click on the HTML output of the Summary Statistics and select **Share > Export**. Save the resulting HTML output in your P:\drive to be uploaded on Canvas. (4 pts)

Problem 4 (8 points): Import Promotion Data

The **promotion** data contains the promotions have given to donors in the past and the current donation status. The names and descriptions of the columns of this data are given in Table 3.

Use an appropriate task to complete the following:

a) Make sure the type of **Control number** and **Status Category 96 NK** characters and anything else is numeric. (2 pts)
b) Change output SAS dataset name to **promotion** and save this file in the WORK library.

c) 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.

Bb Q3) Based on the SAS Enterprise Guide result, summarize your findings into a table: Right-click on the HTML output of the Summary Statistics and select **Share > Export**. Save the resulting HTML output in your P:\drive to be uploaded on Canvas. (3 pts)

d) Use the **One-Way Frequency** task to obtain the frequency counts for each level of StatusCat96NK, StatusCatStarAll, Target_B.

Bb Q4) Based on the SAS Enterprise Guide result, summarize your findings into a table: Right-click on the HTML output of the One-Way Frequency task and select **Share > Export**. Save the resulting HTML output in your P:\drive to be uploaded on Canvas. (3 pts)

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

Problem 5 (40 points): Filter and Sort, Query Builder or PROC SQL

In order to answer the following questions, use Filter and Sort, Query Builder, PROC SQL or any other related SAS EG tasks to get information from the three imported data sets.

a) What are the ID numbers for the youngest male and female donor(s) in the data set **with known age and gender**? (Q5 - Enter your answers on Canvas for Part a.)

Gender	ID
Female	
Male	

b) How many males and females have donated this time for donors **with known gender**? (Note: Target_B takes a value of 1 if a person has made a donation or 0 otherwise.) (Q6 - Enter your answers on Blackboard for Part b.)

Gender	# of donors
Female	
Male	

c) What is the average age for male donors and female donors who made at least three donations in

the past 36 months (**GiftCnt36**) for donors **with a known gender**? (Q7 - Enter your answers on Canvas for Part c.)

Gender	**Avg. Age**
Female	
Male	

d) Create a SAS data set in the 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? (Q8 - Enter your answers on Canvas for Part d.)

# of observations	

e) List the ID numbers for the Donor who has been the donor for the longest time, and how long (in terms of the number of months) have they been a donor? (Q9 - Enter your answers on Blackboard for Part e.)

ID	**# of months**

f) Who (list the ID) made the most recent donations? If multiple made the most recent donation simultaneously, **only list the ID** for who made the largest most recent donation amount. What is the most recent donation amount for that donor? (Q10 - Enter your answers on Canvas for Part f.)

ID	**Amount**

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? (Q11 - Enter your answers on Canvas for Part g.)

# of observations	

Problem 6 (10 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 a 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 (10 points): Stack Columns

a) Create a SAS data set in WORK library named "Promotion_Gender", including Dem- Gender, average PromCnt12, average PromCnt36, average PromCntAll, average Prom- CntCard12, average PromCntCard36, and average PromCntCardAll for males and females only (not for unknown gender donors).

b) Use the Stack Columns task to generate a 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".