DA 6223 Exam 2

Reminder: All Exam solutions must be completed independently.
The SAS Enterprise Guide project (.egp) file must be turned in on Canvas before 11:59 PM CT on Thursday, April 9, 2024 The SAS Enterprise Guide project (.egp) file must be formatted as Firstname_Lastname_ABC123_Exam2.egp.
You should also create separate Process Flows for each problem.

The exam data files can be found on Canvas under the Exam 2 folder.

To complete the tasks, use the SAS EG or PROC SQL or a combination of both. Choose the tool(s) you feel comfortable with as long as they're suitable for each task.

Good luck!

Data Description: The provided data sets include data from a bank. Some of the available variables are defined as follows.

Customer_Demographics

Variable	Description
Education	Customer Education Level
customer_id	ID of customers
Age	Customer Age
Marital	Customer Marital Status
Job	Customer Occupation

Customer_Banking_Info

Variable	Description
customer_id	ID of customers
default	Whether the customer has a default on the account (Y/N)
balance	The current balance in the account
housing	Whether the customer owns a house (Y/N)
loan	Whether the customer carries a loan (Y/N)

Customer_Promo

Variable	Description
customer_id	ID of customers
contact	Preferred contact method
day	Day of Birthday
month	Month of Birthday
duration	Number of days since being a customer
pdays	Number of days since the previous campaign
poutcome	The outcome of the previous campaign

Table 1. Variable Names and Labels

Problem 1 (Total: 30 Points)

Import **customer_demographics** data with the following requirements:

- **Use Query Builder** to save the imported data as **customer_demographics** in the WORK library. (2 Points)
- Sort the data by Age in ascending order. (2 Points)

Based on the imported data, answer the following questions:

a) Create one-way frequency tables for ***Education***, ***Marital***, and ***Job***. (2 Points)

b) Save the results as a PDF file. Go to **Properties > Edit… > Results.** Then check **Customize result formats, styles, and behavior** and check **PDF**. Next, save the PDF on your computer by exporting it. To export: On Problem 1 process flow, right-click on the PDF, then **Share > Export** and select a folder to save the file. (4 Points)

Question 1 on Canvas) Upload the pdf containing the results from One-Way frequency tables on Canvas. (4 points)

c) Create a summary statistics table for *Age*. Save the results as a PDF file. (2 Points)

Question 2 on Canvas) Upload the pdf containing the results from the summary statistics table on Canvas. (4 points)

d) Examine the above outputs: both the one-way frequencies of ***Education***, ***Marital***, and ***Job***, and the summary statistics of **Age**.

Question 3 on Canvas) Do you see any problem/issue in the data for any of the variables? What are the problems/issues in the data? (4 Points)

ANSWER:

e) Create a data set **customer_demographics_recode** named by fixing the issues you identified in Part d. (4 Points)

- For example, for *Education*, recode SECONDARY to **secondary**, etc.
- In addition, for *Age*, recode if *Age* > 100 then set *Age* to **a missing value.**

f) Redo part **a** and a modified version of part **c** to answer the following questions:

Question 4 on Canvas) How many divorced customers are in the data? (2 Points)

ANSWER:

Question 5 on Canvas) What percentage of customers work in an administration position? (2 Points)

ANSWER:

Question 6 on Canvas) What is the average age for the customers who received secondary education? (2 Points)

ANSWER:

Problem 2 (Total: 20 Points)

Import **customer_banking_info** data with the following requirements (2 Points):

- Save the imported data as **customer_banking_info** in the WORK library.
- Change the **names** of the imported variables to match Table 1.
- Set the type of *Customer_ID* to be numeric and format *Balance* appropriately. Based on the imported data, answer the following question:

a) Construct the following table where the dollar amount is the average balance amount. When constructing the table, use an appropriate format for the average balance amount.

Question 7 on Canvas) Please write the correct dollar amounts in the blanks. (8 Points; 2 pts each)

	default	
	no	yes
	Average	Average
loan		
no	$_____	$_____
yes	$_____	$_____

b) Use a pie chart to show the *percentage* of customers who carry and do not carry a loan. Save the chart a PDF. (2 Points)

c) You may customize the chart however you want. But please make sure you do at least the following:
   - Present labels and percentages wherever you want on the chart. (2 Points)

Question 8 on Canvas) Upload the pie chart PDF on Canvas. (4 Points)

Problem 3 (Total: 20 Points)

Import **customer_promo** data with the following requirements (4 Points):

- Save the imported data as **customer_promo** in the WORK library.
- Change the **names** of the imported variables to match Table 1.
- Set the type of *Customer_ID* to be numeric.

Based on the imported data, answer the following questions:

a. Create a data set **customer_promo_recode** named by completing the following tasks: (12 Points)

- For *Month*, recode it into a **numeric variable** such that it takes values 1, 2, 3,…for original values "*jan, feb, mar,…*".
- Create a new variable ***Pdays_group*** based on the original variable **pdays** as

$$Pdays\_group = \begin{cases} -1 & if\ client\ was\ not\ previously\ contacted \\ 1 & if\ 0 \le number\ of\ days\ since\ previous\ campaign \le 273 \\ 2 & if\ number\ of\ days\ since\ previous\ campaign \ge 274 \end{cases}$$

- Create a new variable Last_Promotion_Date based on **pdays**. Assign a **missing value** to Last_Promotion_Date if *pdays* = -1. Format Last_Promotion_Date using DATE9. format. To obtain the last promotion date, you need to obtain a date that is *pdays* earlier than April 9, 2024 (Assume that you are doing this analysis on April 9, 2024).

Question 9 on Canvas) What is the most recent promotion date? Write the date in the format of 01JAN1960. (4 Points)

Problem 4 (Total: 30 Points)

To answer questions for this task, you need to join the imported data together.

a. Create a data set named **Banking** in WORK library by joining **customer_promo_recode**, **customer_banking_info**, and **customer_demographics_recode** together. (6 Points)

b. Assume that the first two digits of the customer ID capture each customer's region information. Following the steps below, construct a Summary Table to count the number of customers by region and contact method.

- Create a data set named **Banking_Region** in the WORK library. (4 Points)
- Three columns should be included in this data: *Customer_ID*, *Region*, and *Contact*. (Note: While *Customer_ID* and *Contact* are already in the **Banking** data, *Region* contains the first two digits of the *Customer_ID*.) (3 Points)
- Define a format named **RegionFmt** for *Region* (defined above) as

$$\text{Format } Region \text{ as } \begin{cases} Region\ A & if\ Region\ is\ 10\ or\ 11 \\ Region\ B & if\ Region\ is\ 12\ or\ 13 \\ Region\ C & if\ Region\ is\ 14 \end{cases}$$

and define a format named ContactFmt for Contact as

$$\text{Format } Contact \text{ as } \begin{cases} Cell\ Phone & if\ Contact\ take\ value\ "cellular" \\ Home\ Phone & if\ Contact\ take\ value\ "telephone" \\ Other & if\ Contact\ take\ value\ "unknown" \end{cases}$$

- Use the data and formats defined above to construct the following summary table. The number represents how many customers are in each region for each contact method. (9 Points)
  Question 10 on Canvas) Based on the summary table you created, please write the correct numbers in the blanks.

	Region A	**Region B**	**Region C**
Cell Phone			
Home Phone			
Other			

c. Based on the merged data **Banking**, create a new variable **Date_of_Birth** to denote each customer's birthday for customers whose *Age* **is not missing**. Find out who is the oldest customer.
Question 11 on Canvas) What is the ID for that customer, and what is his birthday (write your answer in a DDMMMYYYY format, e.g., 01JAN1960)? (4 Points)

d. Based on the merged data **Banking**, **using CASE EXPRESSION** to create a new variable **Balance_Group** to group the customers in the following way:

Balance	**Balance_Group**
$0 \leq Balance < 1000$	Low
$1000 \leq Balance < 3000$	Mid
$3000 \leq Balance$	High
$Balance < 0$	Other

e. **Use a one-way frequency task** on the new variable **Balance_Group** to summarize the number of customers belonging to each group. Save and export the frequency table as a PDF.

Question 12) Upload the PDF on Canvas. (5 Points)