DA 6223 Exam 1

Due: Monday, October 30, 2023, at 11:59 PM

Note: The SAS Enterprise Guide project file needs to be uploaded on Canvas.
Please write your first name, last name, and abc123 in the filename, for example, Isil_Koyuncu_ABC123_EXAM1.egp.
The program needs to run without any running errors. **Any runtime error will result in a 5 points deduction from the total score.**
You should also **create separate Process Flows for each Problem in Part III. Answer Business Scenarios. Failing to do so will result in a 5 points deduction from your grade.**
You must enter your answers to business questions (highlighted in yellow) on Canvas.
Any violation of Academic Conduct (including looking at other people's work, communicating with others about the exam questions while taking the exam, etc.) will result in a 0 on the exam.
Good luck!

I. Conceptual Multiple-Choice Questions

For multiple-choice problems, **only one of the answers is correct. Please select the correct answer on Canvas.**

II. Define a SAS Library

Download the data files shared with you and save them in P:\Exam 1. **Failing to save the files exactly in this path will result in -5 points.**

Define a working library exam1 for the directory where you store the files and limit the access to read-only.

III. Answer Business Scenarios

Business Problem: The provided data sets include data from an automobile insurance company. Some of the available input variables are the car owner's age, number of driving children, marital status, income, motor vehicle record points, number of earlier claims in the past ten years, and whether the driving license was revoked in the past ten years. Three different data sets are provided: **demographics, driving_info,** and **claim_info.** The variables in each data set are summarized in Table 1.

Table 1. Variable Names and Labels

Demographics		Driving_Info		Claim_Info	
Name	**Label**	**Name**	**Label**	**Name**	**Label**
ID	ID number	Driver_ID	Driver ID number	ID	ID number
GENDER	Gender of driver	CAR_USE	Car use (Private/Commercial)	CLAIM_AMOUNT	Total Claim Amount in recent 10yrs
AGE	Age	DISTANCE	Work distance	CLAIM_IND	Claim Flag
STATUS	Marital Status	VEHICLE_AGE	Age of Vehicle	CLAIM_FREQ	Num of claims in recent 10yrs
EDU_LEVEL	Highest Education	VEHICLE_TYPE	Type of Car		
JOB	Occupation of driver	VEHICLE_VAL	Vehicle value		
YOJ	Years on Job	MVR_PTS	Record points		
INCOME	Income	REVOKED	License revoked (10yrs)		
HOUSE_VAL	Housing value				
CHILDREN	Num of children				
CHILD_DRIV	Num of driving children				
AREA	Home/Work Area				

You should use the provided data to **complete the following tasks** and **answer the following questions**.

Problem 1 – Claim_Info

Create a new process flow called Problem 1 and import **Claim_Info** data with the following requirements:

- Save the imported data as Claim_Info in the WORK library.
- Change the names of the imported variables to match Table 1.
- Set the variable type of ID to be 'character' and the length of CLAIM_IND to 8.
- Do not embed the data.

Based on the imported data, answer the following questions:

a) Fill out the following table on Canvas for summary statistics of the claim amount for flagged and non-flagged claims (**round results to 2 decimal places**):

Analysis Variable: Claim_Amount Total Claim Amount in recent 10yrs				
Claim Flag	**Mean**	**Std Dev**	**Minimum**	**Maximum**
0				
1				

Problem 2 – Driving_Info

Create a new process flow called Problem 2 and import **Driving_Info** data with the following requirements:

- Save the imported data as Driving_Info in the WORK library.
- Change the names of the imported variables to match Table 1.
- Set the type of Drive_ID to be 'character.'
- Format the Vehicle_Val variable to be with a dollar sign and an integer.
- Do not embed the data.

Based on the imported data, answer the following questions:

a) How much is the most expensive **private vehicle**? *DO NOT enter the dollar sign on Canvas.*

b) How much is the average value of a **commercial vehicle**? *DO NOT enter the dollar sign on Canvas.*

c) How many drivers who drive a **Sports Car** have their **license revoked**?

d) Which is the most popular (in terms of the number of drivers) known type of vehicle?

Problem 3 – Demographics

Create a new process flow called Problem 3 and import **Demographics** data with the following requirements

- Use Query Builder to save the imported data as Demographics in the WORK library.

Based on the imported data, answer the following questions:

a) How many rows are in the data?

b) What is the **variable length** for the variable EDU_LEVEL?

c) How many drivers are **married** (excluding drivers with "NA" status and missing values)?

d) Which known occupation (that is JOB ≠ 'NONE' and is not missing) has the highest average income?

Problem 4 – Join Tables

Create a new process flow and name it Problem 4. Based on the mutual information (**joined table**) of the provided data, use appropriate tasks to answer the following questions:

a) **For the drivers who filed the claim** (CLAIM_FREQ> 0), what is the average age for each gender (**separately**)? Round the results to 2 decimal places. *(Hint: If SAS displays the calculated column as an integer, modify its format to include two decimal points.)*

b) How many teachers and students (list them separately) have had their driver's license revoked (REVOKED= Yes) in the past?

c) List the ID for the driver with the shortest work distance and driving the Sports Car. If multiple drivers drive the same distance, list the one who filed the largest total claim amount.

Problem 5 – Generate A Summary Table

Use appropriate procedures to generate the following output tables:

We want to create the following table to see the relationship between gender, marital status, and average claim amount. The dollar amounts in the table should refer to the average total claim amount (average of CLAIM_AMOUNT).

a) Enter the values you found in the below table.

Gender	**Married_No**	**Married_Yes**
F	$	$
M	$	$

Note: The Gender column is recoded to have F or M as values, and the Status column is cleaned by removing 'z_' in front of No.

IV. Upload Your SAS EG Project

File response: Please upload the project file (firstname_lastname_abc123_exam1.egp) on Canvas.

Find the grading rubric for the exam below:

Answer the conceptual multiple-choice questions on Canvas (50 Points)
Define a SAS Library (5 Points)
Problem 1

- Import **Claim_Info** data correctly (10 Points)
- Use appropriate tasks or PROC SQL steps to answer the business problems (8 Points)

Problem 2

- Import **Driving_Info** data correctly (10 Points)
- Use appropriate tasks or PROC SQL steps to answer the business problems (16 Points)

Problem 3

- Import **Demographics** data correctly (5 Points)
- Use appropriate tasks or PROC SQL steps to answer the business problems (16 Points)

Problem 4

- Join **Claim_Info, Driving_Info, and Demographics** data correctly (8 Points)
- Use appropriate tasks or PROC SQL steps to answer the business problems (12 Points)

Problem 5

- Use appropriate procedures to generate the table (7+8 Points)

Upload your SAS EG Project (5 Points)