UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

Department of Computer Science & Informatics

CSIA6843

Statistical Analysis System Programming

BASE SAS COMPETENCY PRACTICAL EVALUATION

Assessment: Practical Project

Date Issued: 24 August 2022

Due Date: 25 October 2022

Submission Procedure: BlackBoard

Examiner: Mr. L. Grobbelaar

Moderator: Me. W. Kuyler

Examiner

Moderator

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

Instructions:

1) Please complete all the sections of the project as code segments by making use of the SAS integrated development environment University Edition.
2) Your initials and surname as well your student number should appear at least once at the top of each code section or consecutive code section (as a comment). This is how your work will be identified.
3) Each code section should contain, in commented form, the section number or consecutive section numbers relevant to the code e.g. "Section 1.1" or "Section 1.1 – 1.3". If you present consecutive section numbers as one section of code, please indicate the different section numbers relevant to the code in comments. Code that does not contain this information **will not** be assessed.
4) Where you are required to programmatically create datasets, you are required to include the relevant section number in the form of headers in the output as well as your name and student number in the form of footers. This is mandatory for the output of each dataset created.
5) Ensure the program sections you code successfully execute on different computers.
6) Change the name of the folder containing the input data files to "SASUFS".
7) Save all your code in a secondary folder named "Code" in the "SASUFS" folder.
8) All created datasets, unless specified as temporary, should be considered as being permanent and written to a library named "SASProject".

General:

The mark obtained for this project will account for 30% (thirty percent) of your final mark.

The assignment covers BASE SAS Programming I concepts.

Your assignment is due on the 29th of October 2022.

Submissions should be via BlackBoard in the form of a .zip folder saved as your student number.

Any questions or queries may be raised in the scheduled classes or via E-Mail (LgrobbeL@cut.ac.za).

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

To complete the first, second and third sections, you will need to execute the `cre8data` and `libname` programs accordingly. The data needed to complete sections two and four is contained within the accompanying folder.

Section 1 **[10 marks]**

1. Write a program to display a subset of **orion.employee_addresses** as shown below. The program should sort the observations by **State**, **City**, and **Employee_Name** and then display the sorted observations grouped by **State**. The resulting report should contain 311 observations. (10)

US Employees by State

---------- State=CA ----------

Employee ID	Name	City	Zip Code
120656	Amos, Salley	San Diego	92116
120759	Apr, Nishan	San Diego	92071
121017	Arizmendi, Gilbert	San Diego	91950
121062	Armant, Debra	San Diego	92025
121049	Bataineh, Perrior	San Diego	92126

...

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

Section 2 **[25 marks]**

2.1 Write a PROC PRINT step to display the report below using **orion.sales** as input. Subset the observations and variables to produce the report shown below. Include titles, labels, and formats. The results contain 13 observations. (10)

US Sales Employees
Earning Under $26,000

Employee_ID	First Name	Last Name	Title	Salary	Date Hired
121036	Teresa	Mesley	Sales Rep. I	$25,965	OCT2007
121038	David	Anstey	Sales Rep. I	$25,285	AUG2010
121044	Ray	Abbott	Sales Rep. I	$25,660	AUG1979
...					
121106	James	Hilburger	Sales Rep. I	$25,880	FEB2000
121108	Libby	Levi	Sales Rep. I	$25,930	NOV2010

2.2 Complete the following tasks:

a. Retrieve the starter program **Q2_2.sas** from the supplied data folder.

b. Create a character format named $GENDER that displays gender codes as follows: (4)

F	Female
M	Male
Any other value	Invalid code

c. Create a numeric format named SALRANGE that displays salary ranges as follows: (5)

At least 20,000 but less than 100,000	Below $100,000
At least 100,000 and up to 500,000	$100,000 or more
missing	Missing salary
Any other value	Invalid salary

d. In the PROC PRINT step, apply these two user-defined formats to the **Gender** and **Salary** variables, respectively. Submit the program to produce the following report: (6)

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

Partial PROC PRINT Output:

Salary and Gender Values
for Non-Sales Employees

Obs	Employee_ID	Job_Title	Salary	Gender
1	120101	Director	$100,000 or more	Male
2	120104	Administration Manager	Below $100,000	Female
3	120105	Secretary I	Below $100,000	Female
4	120106	Office Assistant II	Missing salary	Male
5	120107	Office Assistant III	Below $100,000	Female
6	120108	Warehouse Assistant II	Below $100,000	Female
7	120108	Warehouse Assistant I	Below $100,000	Female
8	120110	Warehouse Assistant III	Below $100,000	Male
9	120111	Security Guard II	Below $100,000	Male
10	120112		Below $100,000	Female
11	120113	Security Guard II	Below $100,000	Female
12	120114	Security Manager	Below $100,000	Invalid code
13	120115	Service Assistant I	Invalid salary	Male

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

Section 3 **[20 marks]**

3. Complete the following:

a. Write a DATA step to create **work.delays** using **orion.orders** as input. (2)

b. Create a new variable, **Order_Month**, and set it to the month of **Order_Date**.

Hint: Use the MONTH function. (3)

c. Use a WHERE statement and a subsetting IF statement to select only the observations that meet all of the following conditions:

- **Delivery_Date** values that are more than four days beyond **Order_Date** (1)
- **Employee_ID** values that are equal to 99999999 (1)
- **Order_Month** values occurring in August (1)

d. The new data set should include only **Employee_ID**, **Customer_ID**, **Order_Date**, **Delivery_Date**, and **Order_Month**. (2)

e. Add permanent labels for **Order_Date**, **Delivery_Date**, and **Order_Month** as shown below.

(3)

f. Add permanent formats to display **Order_Date** and **Delivery_Date** as MM/DD/YYYY. (3)

g. Add a PROC CONTENTS step to verify that the labels and formats were stored permanently.

(2)

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Label
2	Customer_ID	Num	8	12.	Customer ID
4	Delivery_Date	Num	8	MMDDYY10.	Date Delivered
1	Employee_ID	Num	8	12.	Employee ID
3	Order_Date	Num	8	MMDDYY10.	Date Ordered
5	Order_Month	Num	8		Month Ordered

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

h. Write a PROC PRINT step to create the report below. Results should contain nine observations. (2)

Obs	Employee_ID	Customer_ID	Order_Date	Delivery_ Date	Order_ Month
1	99999999	70187	08/13/2007	08/18/2007	8
2	99999999	52	08/20/2007	08/26/2007	8
3	99999999	16	08/27/2007	09/04/2007	8
4	99999999	61	08/29/2007	09/03/2007	8
5	99999999	2550	08/10/2008	08/15/2008	8
6	99999999	70201	08/15/2008	08/20/2008	8
7	99999999	9	08/10/2009	08/15/2009	8
8	99999999	71	08/30/2010	09/05/2010	8
9	99999999	70201	08/24/2011	08/29/2011	8

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

Section 4 **[20 marks]**

4.1 Read the instructions below and complete each step accordingly:

a. Open **Q4_1.sas**. Insert a LIBNAME statement to associate the libref **out** with the Excel workbook **employees.xls** in the default data folder. (4)

b. Modify the program so that it creates a spreadsheet named **salesemps** in the **employees.xls** workbook. (6)

c. Submit the SAS program and verify that it created the data set **out.salesemps** with 71 observations and 4 variables as shown in the partial SAS log below.

```
NOTE: 71 records were read from the infile "s:\workshop\newemps.csv".
      The minimum record length was 28.
      The maximum record length was 47.
NOTE: The data set OUT.salesemps has 71 observations and 4 variables.
```

The program will fail if the workbook already exists. Use Windows Explorer to navigate to the data folder and delete **employees.xls**.

4.2 Open **Q4_2.sas**.

a. Write a DATA step to read the delimited in-stream data shown below.

An INFILE statement is required. Use SAS Help or online documentation to explore the use of DATALINES as a file specification in an INFILE statement. (6)

```
120102/Tom/Zhou/M/108,255/Sales Manager/01Jun1993
120103/Wilson/Dawes/M/87,975/Sales Manager/01Jan1978
120261/Harry/Highpoint/M/243,190/Chief Sales Officer/01Aug1991
121143/Louis/Favaron/M/95,090/Senior Sales Manager/01Jul2001
121144/Renee/Capachietti/F/83,505/Sales Manager/01Nov1995
121145/Dennis/Lansberry/M/84,260/Sales Manager/01Apr1980
```

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

b. Generate the report below: (4)

Orion Star Management Team

First	Last	Title	ID	Gender	Salary	HireDate
Tom	Zhou	Sales Manager	120102	M	108255	06/01/1993
Wilson	Dawes	Sales Manager	120103	M	87975	01/01/1978
Harry	Highpoint	Chief Sales Officer	120261	M	243190	08/01/1991
Louis	Favaron	Senior Sales Manager	121143	M	95090	07/01/2001
Renee	Capachietti	Sales Manager	121144	F	83505	11/01/1995
Dennis	Lansberry	Sales Manager	121145	M	84260	04/01/1980

Total: **{75}**

~ END ~