Seneca

Academic Year	**2022 – 2023**		
Semester	☐ Fall	☒ Winter	☐ Summer
Course Code - Name	BAN110		
Instructor	Dr. Razi Iqbal		
Assessment	Assignment 3		

Student ID **176938215**

Student Name **Rongzhao Yi**

Assignment 3

The main purpose of this lab is to get students familiarize with dealing with missing values.

Instructions:

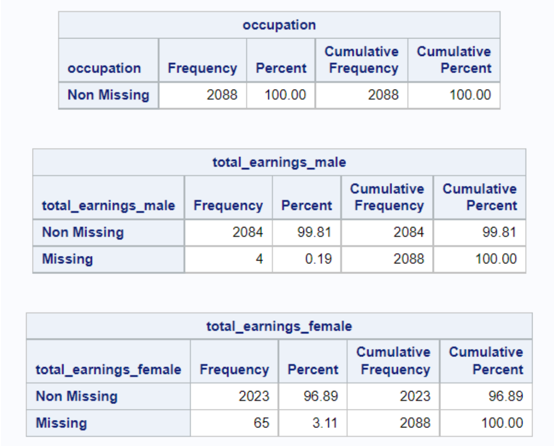
- You are required to submit your answers in this document by pasting your SAS code under the solutions heading below.
- Please do not submit .sas files. Submit only this word document with your code inside it.
- Total Marks for this assignment are 5 marks.
- Students having exactly similar code will get a straight 0.
- You are required to complete these exercises using SAS.
- The deadline for submission of this assignment is March 21, 2023 end of the day.

Question

You are provided with the dataset 'JobsGender.xlsx' file. You are required to perform the following tasks using this dataset in SAS:

- Import the data from the excel sheet in SAS
- Create a dataset in SAS namely JobsGender which would get all the data from Excel sheet.
- Once the data is imported you need to create a new dataset called 'Demography' which would bring only the following columns from the original Orders dataset: Year Occupation Total_Earnings_Male Total_Earnings_Female.

- There are some values in Total_Earnings_Male Total_Earnings_Female columns that are 'NA'. SAS does not recognize these values as missing values. Write a proc format so that all 'NA' values from these two columns are recognized as missing or non-missing as shown in the screenshot below: (You can show occupation in the output as well)



occupation

occupation	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Non Missing	2088	100.00	2088	100.00

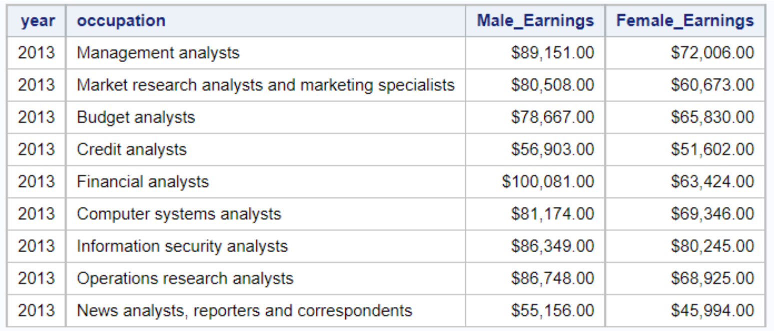
total_earnings_male

total_earnings_male	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Non Missing	2084	99.81	2084	99.81
Missing	4	0.19	2088	100.00

total_earnings_female

total_earnings_female	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Non Missing	2023	96.89	2023	96.89
Missing	65	3.11	2088	100.00

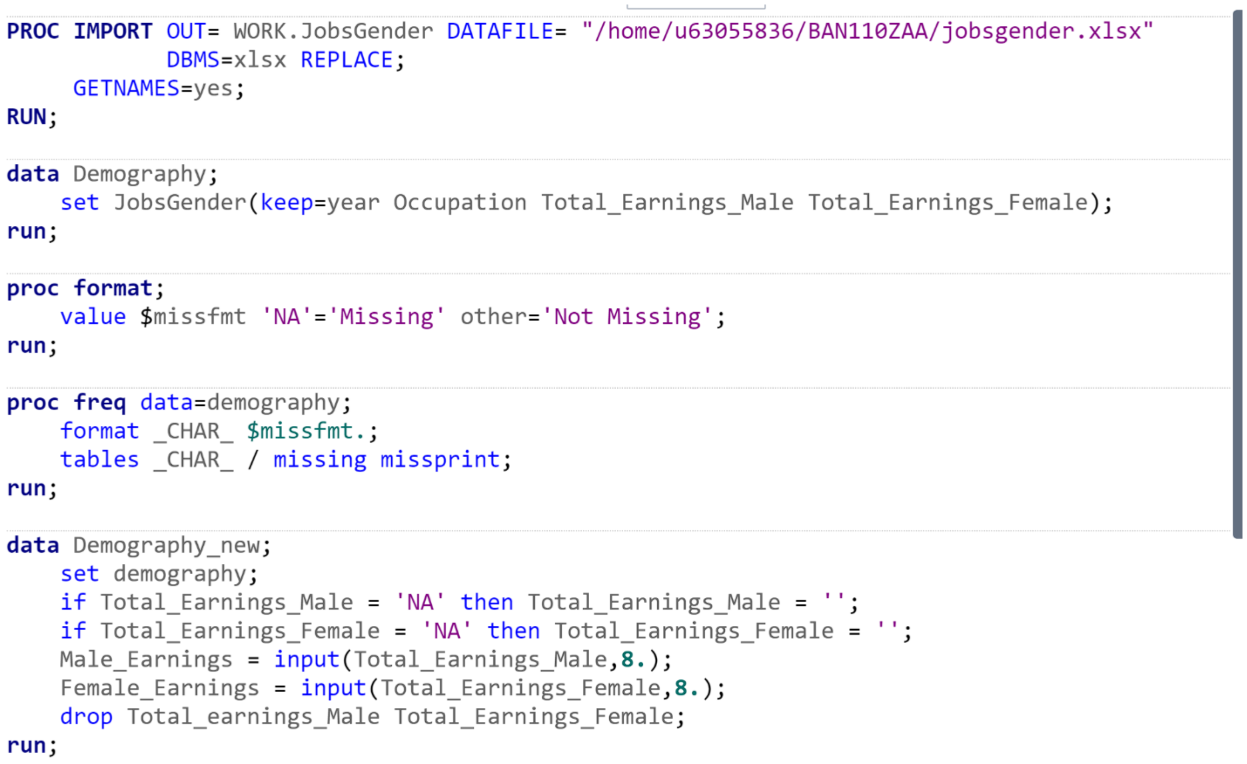
- Once you know how many missing and non-missing values you have, go ahead and replace those 'NA' values by SAS understandable missing values.
- After SAS could recognize those missing values, use Proc Standard to replace all the missing values in those 2 columns by the mean of those columns. (Note that those two columns are character columns and you might want to convert them to numeric columns first.)
- Finally show a list of all the analysts in Year 2013 as shown below. Make sure to format earnings:



year	occupation	Male_Earnings	Female_Earnings
2013	Management analysts	$89,151.00	$72,006.00
2013	Market research analysts and marketing specialists	$80,508.00	$60,673.00
2013	Budget analysts	$78,667.00	$65,830.00
2013	Credit analysts	$56,903.00	$51,602.00
2013	Financial analysts	$100,081.00	$63,424.00
2013	Computer systems analysts	$81,174.00	$69,346.00
2013	Information security analysts	$86,349.00	$80,245.00
2013	Operations research analysts	$86,748.00	$68,925.00
2013	News analysts, reporters and correspondents	$55,156.00	$45,994.00

Make sure to include only the columns shown in the screenshot above.

Solution



```
PROC IMPORT OUT= WORK.JobsGender DATAFILE= "/home/u63055836/BAN110ZAA/jobsgender.xlsx"
            DBMS=xlsx REPLACE;
     GETNAMES=yes;
RUN;

data Demography;
    set JobsGender(keep=year Occupation Total_Earnings_Male Total_Earnings_Female);
run;

proc format;
    value $missfmt 'NA'='Missing' other='Not Missing';
run;

proc freq data=demography;
    format _CHAR_ $missfmt.;
    tables _CHAR_ / missing missprint;
run;

data Demography_new;
    set demography;
    if Total_Earnings_Male = 'NA' then Total_Earnings_Male = '';
    if Total_Earnings_Female = 'NA' then Total_Earnings_Female = '';
    Male_Earnings = input(Total_Earnings_Male,8.);
    Female_Earnings = input(Total_Earnings_Female,8.);
    drop Total_earnings_Male Total_Earnings_Female;
run;
```