Homework 1- SAS Exercise

Instructions:

- You may need to use SAS documentation or The Little SAS book to answer some questions. Alternatively, you can use ChatGPT or other similar tools to get your coding questions answered.
- Feel free to help each other on eLearning discussion boards.

1. The United States Environmental Protection Agency collected data on vehicles with the lowest and highest real-world fuel economy or efficiency in terms of miles per gallon (MPG) by model and year. The data in the file EPA_MPG.csv include each car's model name, weight, drive type, transmission, horsepower, and various MPGs.

Your task is to create a corresponding SAS dataset by reading in this data file.

a. First, examine the raw data file EPA_MPG.csv using Excel.

b. Write a DATA step to read the file into SAS. Make sure that each variable is assigned a unique (and descriptive) name and is of the correct type – character or numeric.

(Hint: use *delimiter = ","* with *infile* command. Use *firstobs* option to skip the header)

c. Create permanent labels for the following variables using the provided descriptions. (HINT: Use the *label* statement).
    - manufacturer: manufacturername
    - weight: weight (lbs)
    - horsepower: horsepower (HP)

d. After the dataset is created, locate the file in the work library through the explorer window in SAS. Double-click on the dataset to view the data. Identify any problems with the SAS data set and explain what is causing the problem. You do not need to propose a solution to solve the problem.

e. Print a report that describes the contents of the data set, including the labels that you have created and other attributes of the variables. (HINT: Use PROC CONTENTS).

2. Now, let us try to import the EPA's MPG data using the import wizard.

a. Examine the raw data file EPA_MPG.csv again and read it into SAS using the IMPORT procedure/wizard as demonstrated during the lecture. Let us name the output file 'EPA_MPG_import2'. Make sure to create a file containing PROC IMPORT statements in the final step of the import wizard.

b. Print the data set (on the results screen). Print a report describing the data set's contents to ensure all the variables are the correct type.

c. Open the raw data file and compare the data values to the output from part b) to make sure that they were read correctly into SAS. In a comment in your report, identify any problems with the SAS data set that cannot be resolved using the IMPORT procedure. Explain what is causing the problem.

d. Read the same raw data file, Pizza.csv, this time using a DATA step (instead of the IMPORT procedure). Be sure to resolve any issues identified above.

(Hint 1: Read sections on common INFORMAT structures in "The Little SAS Book". You do not need to buy the book. There are many websites where you can download the PDF version.)
(Hint 2: You can override the import procedure and modify it by clicking on F4 after opening the PROC IMPORT sas file you generated in part a). Watch the class lecture on SAS I for details.)

e. Create a new dataset with the average values for 3 fuel efficiency variables (i.e., all 3 MPG variables). Specify the number of decimal places to 2 digits. Report the average values in the report.

(Hint: Read pages from "The Little SAS Book" or look up the internet regarding PROC MEANS.)

3. The new management of a local hotel decided to update their recently acquired (and very outdated) property by installing wireless Internet service for their guests. They are also considering updating their billing system because the method used by the previous owner seems faulty. In order to conduct a billing analysis, they would like some calculations about the guests who stayed with them during the first part of February (this was the first month after the change of ownership). The raw data file Hotel.dat contains variables with information on room number, number of guests, check-in month, day, year, check-out month, day, year, use of wireless Internet service, number of days of Internet use, room type, and room rate.

a. Examine the raw data file Hotel.dat and read it into SAS.
b. Create date variables for the check-in and check-out dates, and format them to display as readable dates.

(Hint
Step1: You need to combine three columns into one that looks like a date: for example 2 /7 /2014. You can use "CATX" function.
Step2: If you do step 1, you have a column that has the date. Now you need to let SAS know that this is actually a date. Note that SAS does not realize this by itself, as you have seen in question 1. You can combine "INPUT" function and a proper INFORMAT to let SAS recognize the date.
Step3: Dates are saved as numbers in SAS. To display them as dates in printed output, use a proper FORMAT structure.

c. Create a variable that calculates the subtotal as the room rate times the number of days in the stay, plus a per person rate ($15 per day for each person beyond one guest), plus an Internet service fee ($9.95 for a one-time activation and $6.95 per day of use).

(Hint1: You can subtract dates if they have been stored as SAS dates
Hint2: Since the per person and internet service rates are different for different observations, you can use IF-THEN statement to do the job efficiently.)

d. Create a variable that calculates the grand total as the subtotal plus sales tax at 9.75%. The result should be rounded to two decimal places.

e. View the resulting data set. In a comment in your report, state the value for the grand total for room 532, checked in on Feb. 8th, 2014.