```
/*************************************************************************************/
/* Program Name: STAT 604 HW#10                                                      */
/* Date Created: 10/26/2021                                                          */
/* Author: Jack Rodoni                                                               */
/* Purpose: STAT 604 HW#10                                                           */
/* Date Modified: 10/28/2021                                                         */
/* Location: /home/u59649056/Homeworks/JRodoni_Homework10.sas                        */
/*************************************************************************************/

/* 1.) Housekeeping to clear any titles and footnotes and to turn off the printing of procedure titles*/

TITLE;
FOOTNOTE;
ods noproctitle;

/* 2.) Assign a libref to the mylib folder containing your permanent data sets. If you are going to use */
/*     the professor's data set on SAS Studio, assign a separate library to the Fall2021 folder and add */
/*     access=readonly to the end of the libname statement. Create a fileref to the pdf file for output.*/

libname mylib "/home/u59649056/Homeworks/mylib";
filename pdfCovid "/home/u59649056/Homeworks/mylib/JRodoni_HW10_Output.pdf";

/* 3.) Write a single SAS step that will use the Covid permanent data setas input and create three data sets */
/*     as described in more detail below.Everything in this step must be done as efficiently as possible   */
/*     based on the information you have available.                                                        */

/*      (a) Use a conditional statement that will write out a blue note and the contents of the PDV before */
/*          the set statement on only the first two iterations of the data step.  The message in the note  */
/*          should read "PDV Before Set Statement".                                                        */
/*      (b) The three data sets will only contain rows from the state of Texas.                            */

/*      (c) Since all rows will be from Texas, the state and continent variables are not needed.  The data source name */
/*          is not to be included in the output data sets.  Exclude any column whose name begins with country.() */

/*      (d) The first data set will be a temporary dataset of pre-covid data based on a POSITIVE_CASES_COUNTvalue of 0. */

/*      (e) The second data set will be a permanent data set of covid data where POSITIVE_CASES_COUNT is not 0. */

/*      (f) The third data set will be a permanent data set of all Texas covid data. */

/*      (g) Create a variable of the percent of cases that are fatal by dividing the value of DEATH_COUNT by the value */
/*          of POSITIVE_CASES_COUNT. NOTE: Since the pre-covid data set will not have any values to compute, when the  */
/*          positive cases count is 0, do not process the assignment of this variable or the variable created in the next step. */

/*      (h) Create a character variable containing a fatality group value based on the percent of fatal cases.  About half of */
/*          the observations have a fatality rate of two percent (.02) or less. Give this group a value of Low.  The majority of  */
/*          remaining observations have a value less than 5 percent (.05).  Give this group a value of Medium.  The rest of  */
/*          the observations(with a fatality percent of 5percent or more)will be in the High group. */

/*      (i) Use a conditional statement that will write out a blue note and the contents of the PDV immediately before the */
/*          run statement on only the first iteration of the data step.  The message in the note should read "PDV Before Run Statement". *

data covid_sub1 mylib.covid_sub2 mylib.covid_sub3;
    IF _N_ <= 2 Then put "NOTE- PDV Before Set Statement";
    set mylib.covid;

    where PROVINCE_STATE_NAME = "Texas";

    drop CONTINENT_NAME
         PROVINCE_STATE_NAME
         DATA_SOURCE_NAME
         COUNTRY_SHORT_NAME
         COUNTRY_ALPHA_2_CODE
         COUNTRY_ALPHA_3_CODE;

    IF POSITIVE_CASES_COUNT ^= 0 THEN DEATH_PERCENT = DEATH_COUNT/POSITIVE_CASES_COUNT;

    Length DEATH_GROUP $25;
    IF DEATH_PERCENT <= 0.02 then DEATH_GROUP="Low";
    ELSE IF 0.02 < DEATH_PERCENT < 0.05 then DEATH_GROUP = "Medium";
    ELSE IF DEATH_PERCENT >= 0.05 then DEATH_GROUP = "High";


    IF POSITIVE_CASES_COUNT = 0 Then OUTPUT covid_sub1;
    IF POSITIVE_CASES_COUNT ^= 0 Then OUTPUT mylib.covid_sub2;
    IF POSITIVE_CASES_COUNT ^= ' ' Then OUTPUT mylib.covid_sub3;

IF _N_ = 1 Then put "NOTE- PDV Before Run Statement";
RUN;

/* 4.) Open a PDF destination to receive your output. */
```

```
ods pdf file=pdfCovid;

/* 5.) Write a PROC step that will report a list of data sets in the mylib library without  */
/*     reporting the descriptor portion of the data sets. Supply an appropriate title.   */

proc contents DATA=mylib._All_ NODS;
title1 "Mylib Data";
RUN;

/* 6.) Write another PROC step that will report the descriptor portion of the temporary data set created above.   */
/*     Supply an appropriate title.  */

proc contents data=covid_sub1;
title1 "Covid Subset 1 Table Data";
RUN;

/*  7.) Local media outlets often refer to the area between Baylor University and TAMU as the Brazos  */
/*      Valley.  This area encompasses McLennan, Falls, Robertson,and Brazos counties.  Write a PROC  */
/*      step that will report the data portion of the permanent data set from step 3efor the Brazos  */
/*      Valley counties on a specific day.  Supply a title like Brazos Valley Covid Data as of 01Sep2020  */
/*      but use a macro variable instead of hard coding the date.  Construct the subsetting statement  */
/*      so it can use the same macro variable that is used inthe title.  Ahead of the Title statement  */
/*      and PROC step, write two assignment statements for the macro variable.  The first assignment  */
/*      will supply a value for September1, 2020,and the second a value of September1, 2021.  Execute  */
/*      the first macro assignment statement then execute the Title statement and PROC step.  Execute  */
/*      the second assignment statement along withthe Title statement and PROC step again.  Each execution  */
/*      should produce a page in the output with data from 4 observations.  Be sure you capture the log  */
/*      from each execution. */

%let reportdate=01Sep2020;
%let reportdate=01Sep2021;
TITLE  "Brazos Valley Covid Data as of &reportdate";
proc print data=mylib.covid_sub2;
where COUNTY_NAME in ("McLennan", "Falls", "Robertson", "Brazos") and REPORT_DATE = " &reportdate"d;
RUN;

/* 8.) Close the pdf destination */

ods pdf close;

/* 9.) Use the log and report information contained in your PDFoutput document to find the answers to the questions  */
/*     below and include the answers in acomment section at the bottom of your program file: */
/*     (a) Describe and explain the differences between the three PDVs written to the log. */
/*         The second PDV Before Set Statement shows up after the PDV Before Run Statement */
/*         So it seems that the PDV loops through the whole datastep
/*     (b) How many observations were read from the Covid data set that was created in the previous assignment? */
/*                     153255                                                                                  */
/*     (c) How many observations were written out to the temporary data set? */
/*                     21484                                          */
/*     (d) Which county had the highest death_count on the 2020date? */
/*                     McLennan                                       */
/*     (e) Which county had the highest Highest % Fatal cases on the 2020date? */
/*                     Robertson                                              */
/*     (f) Describe and explain the changes in Percent Fatal Cases from the 2020 date to the 2021 date? */
/* None of the changes are very large. For every county other than McLennan the death rate goes down, however  */
/* Death rate in McLennan goes up. */
```