```
/**************************** SAS Assingment-4 *****************************************/
/*Name:- Alok Kumar Singh                                                              */
/*ID:-   19250990                                                                      */
/**************************************************************************************/

/*Q1. The dataset Dates.csv contains 3000 dates from years 2000 to 2015. Read it into SAS.*/

/*a). Create a new variable which contains the date in format DD/MM/YYYY. */

proc import out= ST662Lib.Dates
datafile= "/courses/d77u30vavpRs0h7u2Ms92/ST662_data/Dates.csv"
dbms=csv replace;
getnames= yes;
run;

data ST662Lib.Dates;
    set ST662Lib.Dates;
    Date1 = mdy(month,day,year);
    format Date1 ddmmyy10.;
run;


/*b). Write code to screen the dataset.*/

proc freq data = ST662Lib.DATES;
    tables day month year date1 / nocum nopercent;
run;

/*4 dates are missing*/

/*c). List any errors identified.*/

proc print data= ST662Lib.Dates;
    where Date1 = . or Year < 2000 or Year > 2015;
run;

/*
Errors identified:-
1) obseravation 685 - the day its showing is 31 and month as 9 which is September and we know
September can have maximum of 30 days.
2) Observation 1169 - Year its showing as 2016 and the dataset can have a data maximum till
2015.
3) Observation 1874 - Year its showing as 1910 and the dataset can have data minimum till 2000.
4) Observation 1913 - the day its showing is 31 and month as 4 which is April and we know
April can have maximum of 30 days.
5) Observation 1942 - which has day as 42 which is not possible for any
month in the calander.
6) Observation 2234 - the day its showing is 31 and month as 11 which is November and we know
November can have maximum of 30 days.
*/


/*Q2. The dataset Bricks.csv contains information on Australian quarterly clay brick production
from 1956 to 1994. Read the data into SAS.*/

/*a). Create a single date variable from the year and quarter variables, and format it so
that it reads as quarterly data. Hint: explore the YYQ function and format ‘yyqs8.’. */

proc import out= ST662Lib.Bricks
datafile= "/courses/d77u30vavpRs0h7u2Ms92/ST662_data/Bricks.csv"
dbms=csv replace;
getnames= yes;
run;

data ST662Lib.Bricks;
    set ST662Lib.Bricks;
    Date1 = YYQ(year,quarter);
    format Date1 yyqs8.;
run;
```

```
/*b). Create a time series plot of the data and comment (briefly - one to two sentences)
on the effects (or not) of season, cycle and trend.*/

title 'Q2 b) Graphs';

proc timeseries data = ST662Lib.Bricks print = seasons plots = decomp;
    /*decomp / mode = mult;*/
    id Date1 interval = qtr accumulate= avg;
    var Bricks;
run;

/*From the graph we can see that there is increasing trend till 1975 and then decrease a bit
due to some error and then again it follows a trend till 1983 and again a decrease due to
error after that it will keep the constant trend till the end. We can see that there is some
seasonality in this graph.
We can also see that the error is fluctuating around 1 from the error graph.*/


/*c). Use an appropriate exponential smoothing method to forecast to the end of 1996.
In your answer, state which type of exponential smoothing you used and why, provide a graph
illustrating the forecasts, and give a table of the forecasts with confidence limits.*/
 title 'Q2 c) Graphs';

proc esm data = ST662Lib.Bricks out = nextyear print = forecasts plot = (forecasts) lead = 12 print= estimates;
    id Date1 interval= qtr;
    forecast Bricks / model = addwinters use= predict transform = log;
run;

/*If the data have no trend or seasonal patterns, then linear is appropriate. If the data
exhibit a linear trend, Holt's methpod is appropriate. But if the data shows trend and seasonal
these methods, on there own cannot handle the problem well. The winters method is based on three
smoothing equations - one for the level, one for trend, and one for seasonality. In this
particual case the data have trend and seasonal so we will use winters method for forecasting
the next year.*/


/*Q3. The dataset LakeHuron.csv contains annual depth measurements at a specific site on
Lake Huron from 1875 to 1972. Read the data into SAS.*/

/*a). Create four new variables that contain the time series depth measurements at lag 1 to 4.*/

proc import out= ST662Lib.LakeHuron
datafile= "/courses/d77u30vavpRs0h7u2Ms92/ST662_data/LakeHuron.csv"
dbms=csv replace;
getnames= yes;
run;


proc expand data = ST662Lib.LakeHuron out = ST662Lib.Lakehuron method = none;
    id year;
    convert depth = x_1dpma / transout= (lag 1);
    convert depth = x_2dpma / transout= (lag 2);
    convert depth = x_3dpma / transout= (lag 3);
    convert depth = x_4dpma / transout= (lag 4);
run;


/*b). Generate scatterplots of depth versus each lag variable.*/

proc sgplot data = ST662Lib.LakeHuron;
    title h = 12pt 'Depth Vs Lag1 Variable';
    Scatter x = depth y = x_1dpma ;
run;

proc sgplot data = ST662Lib.LakeHuron;
    title h = 12pt 'Depth Vs Lag2 Variable';
    Scatter x = depth y = x_2dpma ;
```

```
run;

proc sgplot data = ST662Lib.LakeHuron;
    title h = 12pt 'Depth Vs Lag3 Variable';
    Scatter x = depth y = x_3dpma ;
run;

proc sgplot data = ST662Lib.LakeHuron;
    title h = 12pt 'Depth Vs Lag4 Variable';
    Scatter x = depth y = x_4dpma ;
run;

/*c). Comment on autocorrelation in the data.*/

/*
Autocorrelation:- This term refers to the degree of correlation between the two or more columns
in the the datset. In respect to time series it is correlation between time series and lagged
version.
With respect to this data we can see that
lag1 is highly correlated with Depth as we can see the linear relationship in the plot.
lag2 is moderately correlated with depth as we can see the lineae relationship but we can
also see that the variance is not constant at the end.
lag3 and lag4 are not correlated with depth.
*/
```