```
/* 1. The dataset Dates.csv contains 3000 dates from years 2000 to 2015. Read it into SAS. */


PROC IMPORT DATAFILE='/home/u45187342/my_courses/rafaeldeandrade0/ST662_data/Dates.csv'
    DBMS=CSV
    OUT=dates;
    GETNAMES=YES;
RUN;

proc print data=dates;
run;

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

data date1;
set dates;
new_dates =mdy(month,day,year);
format new_dates  ddmmyy10.;
run;

proc print data=date1;
run;

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

proc freq data=date1;
  tables day month year new_dates /nocum nopercent;
 run;



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

/* There are few errors identified as errors:
total of 6 obseervation as found to be an error
1. Three observation have date as 31 but the month does not contains 31st date in that month
2. one date is wrong which is 42.
3. The observation in the year column is one in 1910 and other is 2016*/

proc print data=date1;
 where year < 2000 or year > 2015 or new_dates=.;
 run;

/* 2. The dataset Bricks.csv contains information on Australian quarterly clay brick production from
 1956 to 1994. Read the data into SAS. */


PROC IMPORT DATAFILE='/home/u45187342/my_courses/rafaeldeandrade0/ST662_data/Bricks.csv'
    DBMS=CSV replace
    OUT=bricks;
    GETNAMES=YES;
RUN;


/* (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.’ */

data bricks1 replace;
      set bricks;
      quarterly_data = yyq( year, quarter);
      format quarterly_data yyqs8.;

/* (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. */

proc timeseries data = bricks1 print = seasons plots = decomp;
    decomp / mode =mult;
    id quarterly_data interval = qtr accumulate = avg;
    var bricks;
```

```
run;

/*comment */
/* from the grpah presented above we could find that there is increase in trend till 1974.
then it follows a trend but there is slight decrease in the year 1983 might be due to error
 and then the trend is same and there is presence of seasonal component in the 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 */

proc esm data = bricks1 out = nextyear
print = forecasts plot = (forecasts)
lead =12
print = estimates;
id  quarterly_data interval = qtr;
forecast  bricks /model=addwinters use = predict transform = log;
run;

/* Data having no trend or seasonal pattern  we use simple smoothing,
data with trend uses holt smoothing
data with trend and seasonal pattern then we use addwinters,
Winter based 3 smoothing equations have
1 for level,
1 for trend,
1 for seasonal component */

/* 3. The dataset LakeHuron.csv contains annual depth measurements at a specific site on Lake Huron
/* from 1875 to 1972. Read the data into SAS. */
PROC IMPORT DATAFILE='/home/u45187342/my_courses/rafaeldeandrade0/ST662_data/LakeHuron.csv'
    DBMS=CSV replace
    OUT=lake;
    GETNAMES=YES;
RUN;

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

proc expand data =Lake out = Lake method =none;
id year;
convert depth=lag1  /transout = (lag 1);
convert depth=lag2  /transout = (lag 2);
convert depth=lag3  /transout = (lag 3);
convert depth=lag4  /transout = (lag 4);
run;

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

proc sgplot data=lake;
scatter x=depth y=lag1;
run;
 proc sgplot data=lake;
scatter x=depth y=lag2 ;
run;
proc sgplot data=lake;
scatter x=depth y=lag3 ;
run;
proc sgplot data=lake;
scatter x=depth y=lag4 ;
run;


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

/* lag1 are highly correlated with depth
   lag2 is moderatly correlated with depth
   lag3 and lag 4 is not correlated with depth  */
```