```sas
/*
 * Program: ST662 - Assignment 4.sas
   Author : Sean O'Riogain (18145426)
   Date   : 6th April 2019
 */

/* Question 1 */

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

proc import datafile='/home/seanoriogain200/ST662/Dates.csv'
    out=dates
    replace
    dbms=csv;
    getnames=yes;
run;

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

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

/*
 * (b) Write code to screen the dataset.
 * (c) List any errors identied.
 */

title 'Question 1c: *** Error: Day is out of range (0 - 31)';
proc print data=dates;
    var day;
    where day < 1 or day > 31;
run;

title 'Question 1c: *** Error: Day is too large for month';
proc print data=dates;
    var day month year;
    where day > 0 and ((day > 28 and month = 2 and mod(year, 4) <> 0) or
                       (day > 29 and month = 2 and mod(year, 4) = 0) or
                       (day > 30 and (month = 4 or month = 6 or month = 9 or month = 11)));
run;

title 'Question 1c: *** Error: Month is out of range (1 - 12)';
proc print data=dates;
    var month;
    where month < 1 or month > 12;
run;

```

```
56 title 'Question 1c: *** Error: Date is not specified';
57 proc print data=dates;
58     var date;
59     where date = .;
60 run;
61
62 title;
63
64 /* Question 2 */
65
66 /*
67  * The dataset Bricks.csv contains information on Australian quarterly clay brick production from
68  * 1956 to 1994. Read the data into SAS.
69  */
70
71 proc import datafile='/home/seanoriogain200/ST662/Bricks.csv'
72     out=bricks
73     dbms=csv
74     replace;
75     getnames=yes;
76 run;
77
78 /*
79  * (a) Create a single date variable from the year and quarter variables, and format it so that it
80  *     reads as quarterly data. Hint: explore the YYQ function and format `yyqs8.'.
81  */
82
83 data bricks;
84     set bricks;
85     YYQ = yyq(year, quarter);
86     format yyq yyqs8.;
87 run;
88
89 /*
90  * (b) Create a time series plot of the data and comment (briefly - one to two sentences) on the
91  *     effects (or not) of season, cycle and trend.
92  */
93
94 proc sgplot data=bricks;
95     title 'Question 2b: Australian Clay Brick Production (1956 - 1994)';
96     series x=yyq y=bricks;
97     xaxis labelattrs=(size=12pt) valueattrs=(size=12pt) label='Year';
98     yaxis labelattrs=(size=12pt) valueattrs=(size=12pt) label='Bricks (Millions)';
99 run;
100
101 title;
102
103 /*
104  * The above time series plot shows a generally increasing trend and this trend is cyclical
105  * with a significant correction taking place around 1983. We can also see that there is a seasonal
106  * component to the series: with production typically increasing during the first 3 quarters of
107  * each year, peaking during the 3rd quarter and falling back during the 4th quarter and the
108  * 1st quarter of the following year, as illustrated by the following sample data extract.
109  */
110
111 title 'Bricks Data - First 20 Observations';
```

```
112 proc sql number outobs=20;
113     select *
114     from bricks
115     order by yyq;
116 quit;
117 
118 /*
119  * (c) Use an appropriate exponential smoothing method to forecast to the end of 1996. In your
120  *     answer, state which type of exponential smoothing you used and why, provide a graph illus-
121  *     trating the forecasts, and give a table of the forecasts with confidence limits.
122  */
123 
124 /* Let's see when the time series data actually ends... */
125 
126 title 'Question 2c: End of time series - last year & quarter';
127 proc sql;
128     select max(yyq) format=yyqs8.
129     from bricks;
130 quit;
131 
132 title;
133 
134 /*
135  * The results of the previous query tells us that the time series data ends with 1994 Q3, which
136  * means that we will need to forecast forward for 9 quarters to get us to the end of 1996.
137  */
138 
139 title 'Question 2c: Lake Huron Depth Forecast to the end of 1996';
140 proc esm data=bricks out=bricks_to_1996 plot=forecasts print=forecasts lead=9;
141     id yyq interval=qtr;
142     forecast bricks /  model=addwinters use=predict transform=log;
143 run;
144 
145 title;
146 
147 /*
148  * In the ESM statements above we use the Winters type of exponential smoothing because of the
149  * presence of both trend and seasonality in the bricks dataset.
150  */
151 
152 /* Question 3 */
153 
154 /*
155  * The dataset LakeHuron.csv contains annual depth measurements at a specic site on Lake Huron
156  * from 1875 to 1972. Read the data into SAS.
157  */
158 
159 proc import datafile='/home/seanoriogain200/ST662/LakeHuron.csv'
160     out=huron
161     dbms=csv
162     replace;
163     getnames=yes;
164 run;
165 
166 /*
167  * (a) Create four new variables that contain the time series depth measurements at lag 1 to 4.
```

```
 */

data huron;
    set huron;
    Depth1 = lag1(depth);
    Depth2 = lag2(depth);
    Depth3 = lag3(depth);
    Depth4 = lag4(depth);
run;

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

proc sgscatter data=huron;
    title "Question 3b: Scatter Plot Matrix for Depth versus its 4 Lag Variables";
    matrix depth depth1 depth2 depth3 depth4 / group=year;
run;

title;

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

ods text="Question 3c: The matrix scatter plots printed above indicate that the level of correlation between the Depth variable and its lagged versions decreases as the lag

/*
 * The matrix scatter plots printed above indicate that the level of correlation between the
 * Depth variable and its lagged versions decreases as the lag level increases.
```