A client of yours sent you a data set that contains the customer information from a loyalty card program.  
  
The data set contains tons of information that your client can explore.  
  
However, the data set must be cleaned up before the analysis can be done.  
  
Your boss is currently on vacation, and your client needs this done by tomorrow.   
  
It is the perfect time to show your boss's boss you can totally step up and do the job.  
  
  
**Task 1**  
  
Run the code in the yellow box below to get the **CUSTOMER** data set on SAS Studio. 

The **CUSTOMER** data set contains the following 13 variables:

* **CUSTID:** Customer ID
* **FIRST:** First Name
* **LAST:** Last Name
* **GENDER:** Gender
* **DOB:** Date of Birth
* **POSTAL:**Postal Code (Canada)
* **EDU:** Level of Education
* **OCCUP:** Occupation
* **INCOME:** Income
* **STATUS:** Marital Status
* **NUMKIDS:** Number of Kids
* **SPEND:** Spending in the year past
* **DOS:** Date of Study

Check the followings after you create the **CUSTOMER** data set on SAS Studio:

1. There are 2000 observations in the data set
2. All 13 variables are captured in the **CUSTOMER** data set

Task -2

**Create an Error Report**

Your client requires you to send them an error report that contains all of the potential errors from the data set.  
  
You must perform the following checks on the data set, and create an error report if any error(s) are found.  
  
Your client also doesn't have access to SAS and, as a result, you must export the error report to an Excel spreadsheet.  
  
  
**Task 2a**  
  
Perform the following checks on the**CUSTOMER** data set:  
  
1. Range Check  
  
Range check is commonly done on numeric variables. It checks for the numeric value(s) that is out of the logical range.  
  
Perform range checks on the following variables:

* **INCOME**  
  Identify incomes greater than $500,000 or less than zero
* **SPEND**  
  Identify Spending greater than 3 times of Income or less than zero
* **AGE**   
  Identify Age > 140 or less than zero. Age is calculated as (DOS-DOB)/365.25

2. Invalid Character Check  
  
Invalid Character Check looks for unwanted characters from a character variable.  
  
Perform Invalid Character checks on the following variables:

* **CUSTID**  
  Customer ID should be only 8-character long. It should contain only numbers.
* **FISRT/LAST**  
  First and Last name should contain only letters. No special character(s) are allowed.
* **OCCUP**   
  Occupation should consist of only letters. No special character(s) are allowed.

3. Category Value Check  
  
Category Character Check ensures all categorical and ordinal variables contain only the valid value.   
  
Perform Category Value checks on the following variables:

* **GENDER**  
  Gender can only be either Male or Female. Any other value is considered an error.
* **EDU**  
  Education is actually a coded variable. It should contain only the value of (1, 2, 3, 4).   
  Note: the properties of a coded variable will be explained in details in the next module.
* **STATUS**  
  Marital Status can only be Married, Single or Divorced.

**Task 2b**  
  
Create a data set that contains all of the errors found from Task 2a.   
  
The data set should contain the following variables:

* CustID (e.g. 80050123)
* Variable (e.g. FIRST)
* Value (e.g. %##%)
* Comment (e.g. the customer’s first name contains special characters.)

Export the data set to an Excel spreadsheet and save it as **Error Report**.   
  
Submit the final program in the link below. You will receive the model answer within 1-2 hours.  
  
  
Purpose of the project  
  
When cleaning up data in a practical business environment, you are often required to create an error report on the data set. The error report should list out in details the errors associated with each variable. The report will then be sent to the database administrator (or the data source) for data validation purposes.  
  
Since SAS is not commonly used outside of the statistical analysis environment, you likely need to export the report into a more common business file such as Excel spreadsheet.