|  |  |
| --- | --- |
| Student ID: | D10126532 |
| Student Name: | John Warde |
| Course Code: | DT230B |

Below are descriptions/notes on the approaches I took in the execution of this assignment.

1. Used github for Source Code Management <https://github.com/johnwarde/sasassignment>  
   git config --global user.name "John Warde"  
   git config --global [john.warde@gmail.com](mailto:john.warde@gmail.com)  
   git push origin master
2. At the start I used Powershell to create smaller versions of all the file s to reduce the run times in the early stages of getting the data into the SAS system, in the correct format:   
     
   Get-Content -TotalCount 40 bills.csv | Out-File bills-small.csv  
   Get-Content -TotalCount 40 calls.csv | Out-File calls-small.csv  
   Get-Content -TotalCount 40 callSummaries.csv | Out-File callSummaries-small.csv  
   Get-Content -TotalCount 40 demographics.csv | Out-File demographics-small.csv
3. For all arithmetic division, used the divide() function to cater for the division by zero scenarios.
4. Used linux command (via git bash) “less” to view data any number of lines in, for fixing import errors.
5. Used the frequency reports to detect anomalies in the imported CSV data. Then added statements to convert these anomalies and make the data consistent. e.g. some true/false values had ‘t’ and ‘f’ values these were converted to ‘true’ and ‘false’.
6. Converted “unknown” values in the regionType column to the “missing value” to correctly report on frequency. Also changes “0” values in serviceArea column to “missing value”.
7. In frequency reports, I included the ‘missing’ option to include missing values in the percentage calculations for a more accurate reflection of the occupation, regionType and newCellUser.