CLEANING LOG

Daily Activity:

* Filtered data to determine the number of users within the data set (33 users) and the period of activity for each user (31 days: 12/04/2016 – 12/05/2016)
* Sorted data by date – Oldest to Newest.
* All data was formatted into their appropriate format as data was not consistent and incorrectly formatted throughout  
  - There was an issue with changing the date format, using ‘Text to Column’, I addressed the current format and this resolved the issue, the dates were changed into the format DD/MM/YYYY, numerical values into the Number format and of 1 decimal place.
* Ensured all ID numbers were of correct length using LEN.
* No duplicates were discovered.

Sleep Day:

* Split SleepDay column containing date and time into two separate columns of more appropriate column names of ‘Date’ and ‘Time’ using ‘Text to Columns button’ in Data tab.
* Filtered data to determine the number of users within the data set, there were less users a part of this dataset (24 users of the 33 users) and the period of activity for each user (31 days: 12/04/2016 – 12/05/2016)
* Sorted data by date – Oldest to Newest.
* There was an issue with changing the date format, using ‘Text to Column’, I addressed the current format, and this resolved the issue, the dates were changed into the format DD/MM/YYYY, numerical values into the Number format and of 0 decimal places, Time changed to Time format 12:00:00 AM
* Ensured all ID numbers were of correct length using LEN.
* Removed three duplicate rows of data using Concatenate function to produce unique identifiers, used Conditional formatting on UID and filtering to view only duplicates which were then removed from the dataset.

Weight Log info

* Ensured all ID numbers were of correct length using LEN and as this was a larger dataset, I used the fill-down option to apply the formula for cells of ID.
* Split column containing date and time into two separate columns of more appropriate column names of ‘Date’ and ‘Time’ using ‘Text to Columns button’ in Data tab.
* Dates were changed into the format DD/MM/YYYY, numerical values into the Number format and of 1 decimal places. Time changed to Time format 13:00:00
* Filtered data to determine the number of users within the data set, there were less users a part of this dataset (8 users of the 33 users) and the period of activity for each user (31 days: 12/04/2016 – 12/05/2016)
* Sorted data by date – Oldest to Newest.
* No duplicates found
* Overall 0-2dp for Numerical values on all data sets