Bellabeat Data Cleaning and Profiling

* Renamed tables to standardize table names and maintain schema consistency.
* During importation, all columns were designated as varchar(50). Altered some of the columns to appropriate and applicable datatypes.
  + Updated hourly\_intensity table
  + Updated daily\_sleep table
  + Updated minute\_sleep table
  + Updated daily\_activity table
  + Updated weight\_log table
  + Updated minute\_intensity\_narrow table
  + Updated minute\_mets\_narrow table
* Created minute\_activity table similar to daily\_activity by combining data from more focused tables that use the minute time frame.
* Created hourly\_activity table similar to daily\_activity by combining data from more focused tables that use the hourly time frame.
* Found the count of users in the minute\_activity table, the earliest and latest days in the study, as well as the calendar math from start to finish.
* Found the count of users in the hourly\_activity table, the earliest and latest days in the study, as well as the calendar math from start to finish.
* Found the count of users in the daily\_activity table, the earliest and latest days in the study, as well as the calendar math from start to finish.
* Found the count of users in the daily\_sleep table, the earliest and latest days in the study, as well as the calendar math from start to finish.
* Found the count of users in the hourly\_intensity table, the earliest and latest days in the study, as well as the calendar math from start to finish.
* Found the count of users in the minute\_sleep table, the earliest and latest days in the study, as well as the calendar math from start to finish.
* Found the count of users in the weight\_log table, the earliest and latest days in the study, as well as the calendar math from start to finish.
* Checked the level of record saturation per user Id by comparing the count of TotalSleepRecords per Id in the daily\_sleep table with the count of Calorie records per Id in the daily\_activity table.
* Checked for duplicate rows in minute\_activity table. None found.
* Checked for duplicate rows in hourly\_activity table. None found.
* Checked for duplicate rows in daily\_activity table. None found.
* Checked for duplicate rows in daily\_sleep table. Found 3 duplicate records.
* Created a new sleepDay2 table and populating it with all distinct values from daily\_sleep.
* Dropped the original sleepDay table.
* Renamed new sleepDay2 table to sleepDay and completed the replacement of the table with no duplicates as verified by an additional check.
* Checked the hourly\_intensity table for duplicate records. None found.
* Checked the minute\_sleep table for duplicate records. Found 543 duplicate records.
* Completed process similar to sleep\_day table to replace minute\_sleep table.
* Checked the weight\_log table for duplicate records. None found.
* Checked daily\_activity table for NULL’s and 0’s
* Checked all tables for consistent record character length.
* Checked all tables for NULL values.
* Checked all tables for record outliers.