**BELLABEAT FITNESS COMPANY**

**CHANGE LOG FOR CLEANING AND FORMATTING IN EXCEL**

**REFERENCE FOR CLEANING DATA**

* ADDED – New feature introduced
* CHANGED – Changes in existing functionality
* DEPRECATED – Features about to be removed
* REMOVED – Features that have been removed
* FIXED – Bug fixes
* SECURITY – Lowering vulnerabilities

**VERSIONS**

Contents

[**VERSION 1.0.0 (30/06/2023)** 1](#_Toc180725599)

[**VERSION 1.1.0 (08/09/2023)** 5](#_Toc180725600)

[**VERSION 1.2.0 (01/10/2023)** 6](#_Toc180725601)

# **VERSION 1.0.0 (30/06/2023)**

**DATA CLEANING**

Setting up and streamlining datasets gotten from FitBit Fitness Tracker data: eliminating redundant datasets and aggregating the datasets I need to query and analyze information from which are:

* DailyActivity\_merged
* HourlyCaloriess\_merged
* HourlyIntenities\_merged
* HourlySteps\_merged
* SleepDay\_merged
* WeightLogInfo\_merged

**DAILYACTIVITY\_MERGED**

* **ADDED**
* MERGING DATA
  + Created duplicate **DailyActivities\_merged** dataset as sheet 2
  + Imported **sleepDay\_merged1** dataset as sheet 3
  + Imported **hourlyCalories\_merged** dataset as sheet 4
  + Imported **hourlyIntensities\_merged1** dataset as sheet 5
  + Imported **hourlySteps\_merged1** dataset as sheet 6
  + Imported **weightLogInfo\_merged1** dataset as sheet 7

**\*\*The above was done for analysis within excel**

**\*\*I created individual worksheets for each of the above selected datasets for the purpose of being imported into other analytics software for further analysis\*\***

**Imported all the required datasets key to my analysis from separate excel worksheets into one worksheet**

* **CHANGED**
* DATA TRANSFORMATION
  + Changed data format under **id** column from a general format to a number format.
  + Formatted **ActivityDate** column from a general to date style format
  + Changed data format under the **TotalSteps** column from a general to number style format
  + Changed data format under **TotalDistance** column from a general format to a number format and also reduced it to 2 decimal places for legibility
  + Changed data format under **TrackerDistance** column from a general form to number format and reduced it to a max of 2 decimal places to promote legibility
  + Formatted **VeryActiveDistance** column from a general to number format, reducing it to a max of 2 decimal places for legibility
  + Formatted **ModeratelyActiveDistance** column from a general to number format, reducing it to a max of 2 decimal places for legibility
  + Formatted **LightActiveDistance** column from a general to number format, reducing it to a max of 2 decimal places for legibility
  + Formatted **SedentaryActiveDistance** column from a general to number format, reducing it to a max of 2 decimal places for legibility
  + Formatted **LoggedActivitiesDistance** column from a general to number format, reducing it to a max of 2 decimal places for legibility
  + Formatted **VeryActiveMinutes** column from a general to number style format
  + Formatted **FairlyActiveMinutes** column from a general to number style format
  + Formatted **LightlyActiveMinutes** column from a general to number style format
  + Formatted **SedentryMinutes** column from a general to number style format
  + Formatted **Calories** column from a general to number style format
* **INCONSISTENT DATA**
  + A contradiction exists between what the dataset is said to contain and what it actually contains, as there are **33** **users** in the master sheet as against the **30** it was supposed to contain.

**HOURLYCALORIES\_MERGED**

* **CHANGED**
* DATA TRANSFORMATION
  + Formatted **ActivityHour** column from general to a datetime style format
  + Formatted **ID** column from general to a number style format
  + Formatted **Calories** column from a general to number style format

**HOURLYINTENSITIES\_MERGED**

* **CHANGED**
* DATA TRANSFORMATION
  + Formatted **ID** column from a general to number style format
  + Formatted **ActivityHour** column from general to a datetime style format
  + Formatted **AverageIntensity** column from a general to number format, reducing it to a max of 2 decimal places for legibility
  + Formatted **TotalIntensity** column from general to a number style format

**HOURLYSTEPS\_MERGED**

* **CHANGED**
* DATA TRANSFORMATION
  + Formatted **ID** column from general to a number style format
  + Formatted **ActivityHour** column from general to a datetime style format
  + Formatted **StepTotal** column from general to a number style format
* **ADDED**
  + Added column named “HourOfDay” extracting the hour format (h:mm AM/PM) from ActivityHour column to be used in later analysis.

**SLEEPDAY\_MERGED**

* **REMOVED**
* DUPLICATES
  + Removed 3 duplicated data leaving 410 unique values in **SleepDay \_merged** dataset
* **CHANGED**
* DATA TRANSFORMATION
  + Transformed datetime to date format under **SleepDay** column
  + Formatted **ID** column from general to a number style format
  + Formatted **TotalStepsRecords** column from general to a number style format
  + Formatted **TotalMinutesAsleep** column from general to a number style format
  + Formatted **TotalTimeInBed** column from general to a number style format

**WEIGHTLOGINFO\_MERGED**

* **CHANGES**
* **DATA TRANSFROMATION** 
  + Formatted **Date** column from general to a datetime style format
  + Formatted **WeightKg** column from a general to number format, reducing it to a max of 2 decimal places for uniformity
  + Formatted **WeightPounds** column from a general to number format, reducing it to a max of 2 decimal places for uniformity
  + Formatted **BMI** column from a general to number format, adjusting it to a max of 2 decimal places for uniformity
  + Formatted **ID** column from general to a number style format
  + Formatted **Fat** column from general to a number style format
  + Formatted **IsManualReport** column from general to a text style format

# **VERSION 1.1.0 (08/09/2023)**

**DATA PREPARATION**

I am preparing the datasets selected for further analysis to determine if they can answer the questions indicated in the business task.

**DAILYACTIVITIES\_MERGED**

* **SORTING AND FILTERING**
  + Sorted by **ActivityDate** column from the oldest to the most recent
  + Then by **TotalSteps** from Largest to smallest
  + Filtered out the zero (0) values in the **TotalSteps** column
  + Filtered out the zero (0) values in the **TotalDistance** column
  + **Added an extra column named “day of the week” where I used the TEXT() function to extract the day of the week of each individual date for later analysis**

**SLEEPDAY\_MERGED**

* **SORTING AND FILTERING**
  + Sorted by **sleepday** column from the oldest to the most recent
  + And then by **TotalMinutesAsleep** from the largest to the smallest

**HOURLYCALORIES\_MERGED**

* **SORTING AND FILTERING**
  + Sorted by **ActivityHour** from the oldest to the most recent

**HOURLYINTENSITIES\_MERGED**

* **SORTUNG AND FILTERING**
  + Sorted by **ActivityHour** column
  + Sorted by **Totalintensity** column
  + Filtered out all the zero (0) values from the **TotalIntensity** column

**HOURLYSTEPS\_MERGED**

* **SORTING AND FILTERING**
  + Sorted by **ActivityHour** column from the oldest to the most recent
  + Then by **StepsTotal** column from the Largest to the smallest

**WEIGHTLOGINFO\_MERGED**

* **SORTING AND FILTERING**
  + Sorted by **Date** column from the oldest to the most recent
  + Then by **WeightKg** from largest to smallest

# **VERSION 1.2.0 (01/10/2023)**

**ADDED**

* **DAILY\_ACTIVITY\_MERGED**
  + Added two columns namely **ActivityType** and **RestType** with the intention of simplifying my analysis process in gaining insights. Where activitytype groups data based on the total number of steps users take in a day while rest type categorizes users based on their sleep quality in hours.

**Exported cleaned data to Tableau for visualization** [**CLICK HERE**](https://public.tableau.com/app/profile/mgbecheta.paschal/viz/BellabeatUserAnalysis/BELLABEATSTEPSACTIVITYDASHBOARD#1) **to view harts**

**For analysis report** [**CLICK HERE**](https://github.com/paschaldubem/Bellabeat-Market-Analysis) **to view in GitHub repository**