Data Cleaning Documentation

Date: May 07, 2025

Project: Bellabeat-Fitbit Case Study

Prepared by: [Ahmad Abubakar]

# 1. Overview

This document outlines the data cleaning process conducted on the Fitbit datasets used in the Bellabeat Google Data Analytics Capstone Case Study. The goal was to prepare the data for analysis by handling missing values, duplicates, inconsistent formats, and irrelevant entries.

# 2. Datasets Used

- dailyActivity\_merged.csv  
- sleepDay\_merged.csv

# 3. Data Cleaning Steps

## 3.1 Remove Duplicates

Removed duplicate rows in each dataset using Excel's 'Remove Duplicates' feature or the UNIQUE function.

## 3.2 Date Format Standardization

Standardized all date columns (e.g., 'ActivityDate', 'SleepDay') to the 'mm-dd- yyyy' format.

## 3.3 Merge Datasets

Merged dailyActivity\_merged with sleepDay\_merged using VLOOKUP to align sleep data with corresponding activity data.

## 3.4 Filter Out Zero Values

Filtered out rows where all major metrics (e.g., TotalSteps, Calories, TotalMinutesAsleep) were zero indicating no device usage or invalid entries.

## 3.5 Missing Data Handling

Handled missing data using IFERROR in Excel or by leaving unmatched cells blank after merging.

## 3.6 Data Type Consistency

Ensured numeric columns were properly formatted and not stored as text.

# 4. Notes and Assumptions

- Users did not consistently track all types of data.  
- Zeros were treated as missing or non-tracking days.  
- The merging key assumes daily uniqueness per user.

# 5. Tools Used

Microsoft Excel for data exploration, cleaning, merging, Analysis, and visualization preparation.