🔧 Data Cleaning & Processing Summary

Tools Used:

- Power BI (Power Query)
- Excel

- Steps Taken:

1. Removed Duplicates
   - Identified and removed duplicate rows from datasets such as Daily Activity and Sleep Data using Power Query.
2. Handled Missing Values
   - Columns with more than 90% null values (e.g., `Fat`) were excluded from analysis.
   - Minor nulls were removed or imputed when necessary.
3. Rename columns
   - Modify column names to improve readability by replacing spaces with underscores (_).
4. Data Type Corrections
   - Converted columns like `Date` and `Time` to appropriate formats.
   - Ensured numerical values are in correct data type (e.g., steps, calories, distances).
5. Filtered Invalid Records
   - Removed records with total steps = 0 or unrealistic sleep duration.
   - Excluded users with extremely short tracking periods.
6. Created New Calculated Columns
   - `Sleep Efficiency = MinutesAsleep ÷ TimeInBed`
   - Categorized users by:
     - Step Level: Sedentary, Low Active, Active, Highly Active
     - BMI Category: Underweight, Normal, Overweight, Obese
7. Data Merging
   - Merged datasets on `ID + Date` using Power BI to relate sleep, activity, and calories for each user per day.
8. Unpivoting
   - Used unpivot for minute-level data like METs and steps to restructure narrow tables for visualization.
9. To prepare the data for analysis, I identified tables with identical structures across different monthly files from 3/12/2016–4/11/2016 and 4/12/2016–5/12/2016). These tables contained the same column names and formats but were split across separate files. I performed a data merge (append) for each type of table (e.g., daily activity, sleep data, METs) to combine them into unified datasets. This allowed for a complete and continuous view of user data across both time periods