

Process

We will process three dataset which is used in this analysis

- dailyActivity_merged
- sleepDay_merged
- hourlySteps_merged

Before making changes on these datasets, make a copy of it in a separate folder. And all data processing is done in MS Excel.

dailyActivity_merged

1. Consistent date format
STEP 1: new_col: `=TEXT(date, "d.mm.yy")`
STEP 2: another_new_col: `=DATE(2016, LEFT(new_col, 1), MID(new_col, 3,2))`
STEP 3: Apply auto-fill for both of the above steps
STEP 4: Copy & Paste special > Paste Values, in date_col
STEP 5: Format date_col as Date
2. Duplicate data
STEP 1: In Data ribbon, click Remove Duplicates
3. Data Errors
0 Calories burnt of 4 observations of 4 different Ids with same
 - SedentaryMinute = 1440
 - TotalSteps = 0

By looking at each of 0 calories observations, it seems like a data error where the device is not used and still input 0 values for each attribute except SedentaryMinute. So delete these rows.

- STEP 1: Filter records with 0 calories.
- STEP 2: Filter any one Id from the filtered records.
- STEP 3: Remove the bottom row with 0 calories observation.
- STEP 4: Remove filters from calories observation.
- STEP 5: Repeat the above steps for other observations.

sleepDay_merged

1. Consistent date format
Apply all steps used in `dailyActivity_merged` dataset in the date format section.
2. Duplicate data
STEP 1: In Data ribbon, click Remove Duplicates

hourlySteps_merged

1. Consistent datetime format

STEP 1: Split datetime into 3 columns using feature Text to Columns in Data ribbon.

Now datetime is split into three columns (date, time, period).

STEP 2: Shift period column into another blank space

STEP 3: Concatenate the time and period, then use text function to change the format into 24hr by:

`=TEXT(CONCATENATE(TEXT(time, "hh:mm:ss"), " ", period), "hh:mm:ss")``

This should be next to the time column, not the period column, so that we can easily apply the Auto-fill feature of Excel.

STEP 4: Copy and Paste Special > Paste Values in Datetime column.

STEP 5: Format datetime as Time

2. Duplicate data

STEP 1: In Data ribbon, click Remove Duplicates

----- X -----