These are following datasets that needs to be checked for data quality, consistency and are to be cleaned, structured if required.

1. Daily\_activity.csv
2. Hourly\_merged.csv
3. Minutes\_merged.csv
4. SleepDay.csv

For Daily Activity:

1. Checked Date data type in google sheets.
2. Added two new columns Weekday (Day name) and weekday\_num (day of the week)
3. Deleted Column Tracker\_distance as it is same as Total\_distance
4. Deleted Column LoggedActivityDistance as it is mostly 0 for all the records excepts for few and it does not seem to help with any insight
5. Checking for the consistency of the data
6. Changed the column name of FairlyActiveMinutes to ModeratelyActiveMinutes to resemble with ModeratelyActiveDistance.
7. Added total\_activity\_minutes column which adds up all the recorded activity time
8. Added unrecorded\_minutes and unrecorded\_hours for the activity time that is not recorded by the watch meaning the wearable device was not used.
9. Changed the column name of “Calories” to “Calories\_burned”.
10. Merged SleepDay.csv data of TotalMinutesAsleep, TotalSleepRecorded, TotalTimeInBed to Daily Activity data (Merged in BigQuery) based on Ids and Date.

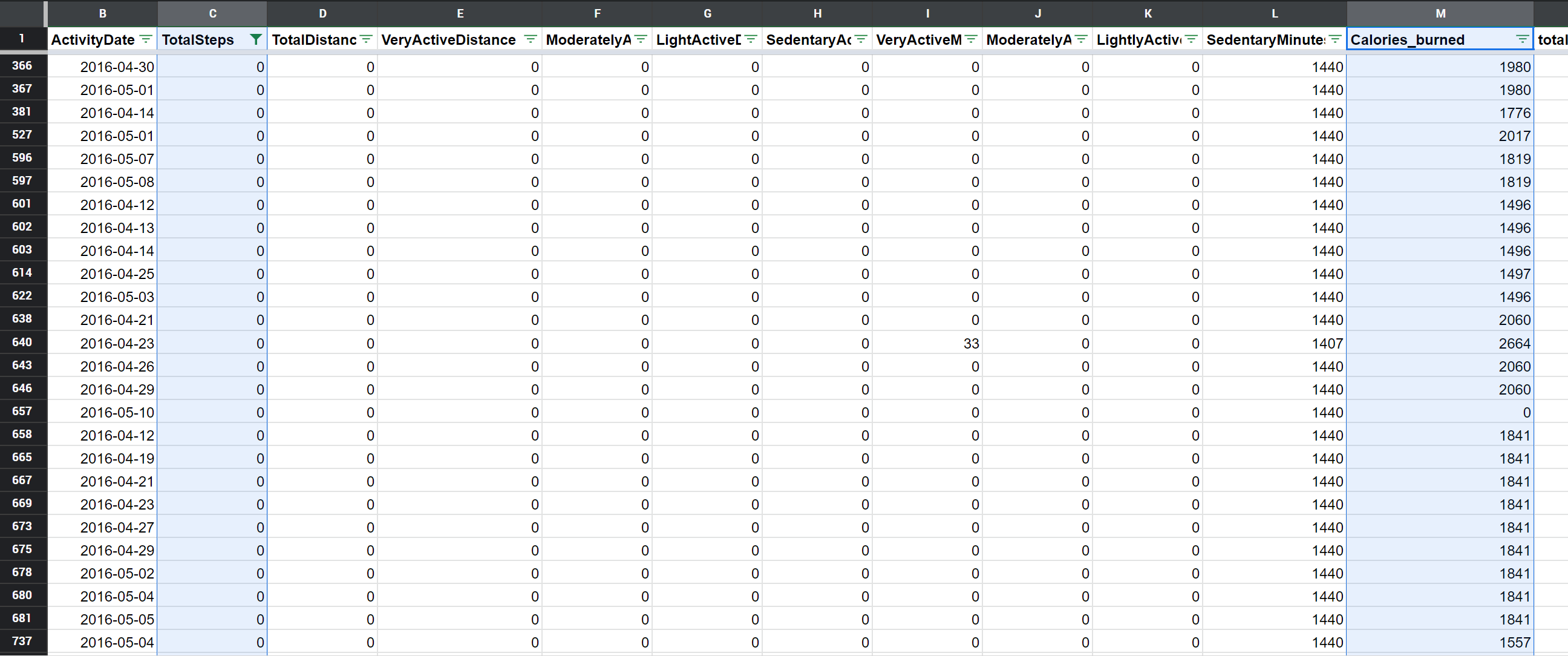
SleepDay:

1. Some sleep data is very random and does not align with the average minutes of sleep time in bed. This is because some participants use devices that do not measure sleep like Fitbit Zip or Fitbit one. Therefore, some data cleaning is done on the minutes of these users.
2. Zero down the minutes of sleep for those Id whose records only show few days of sleep.

Graphical user interface

Description automatically generated

1. When we look at the entries with TotalSteps as 0, we some calories burned values in the table as shown in the below picture. There is no track record of other activity minutes, which means that the device might not have been used during that day and it was not able to record any activity. We will 0 the calories burned for these which might not be true in reality because it is just not recorded.



Hourly Merged:

1. Created an hour column to get the hours from the ActivityHour and created column in Power BI describing hours in AM and PM.
2. Another column called Intensity group is created to group the users and days when they were able to perform activities with different intensity levels. These were grouped in “Light”, “moderate”, and “Vigorous” intensity levels. [[Reference](https://www.omicsonline.org/articles-images/2157-7595-6-220-t003.html)]

Minute\_MET:

1. This table has been converted to show average hourly METs and is merged with hourly table.

For heartrate\_seconds:

1. Changed the data type of Date to Date Time in google sheets
2. The data being too large no other functions can be performed in the Google sheets. It will be checked further in BigQuery