-- first let's explore the number of users in each table

SELECT COUNT(DISTINCT(Id)) AS UniqueNoOfUsers

FROM `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_activity\_merged`; --34 users exist here

SELECT COUNT(DISTINCT(Id)) AS UniqueNoOfUsers

FROM `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_sleep\_merged`; --24 users exist here

SELECT COUNT(DISTINCT(Id)) AS UniqueNoOfUsers

FROM `da-capstone-bellabeat-399911.Bellabeat\_case\_files.hourly\_steps\_merged`; --33 users exist here

SELECT COUNT(DISTINCT(Id)) AS UniqueNoOfUsers

FROM `da-capstone-bellabeat-399911.Bellabeat\_case\_files.weight\_merged`; --8 users exist here

--although the dataset contained 30 users who consented to the usage of the personal data, we find these tables either have less or more than 30 users.

--add columns for the day of the week, sedentary hours, and total active minutes to daily\_activity\_merged csv

ALTER TABLE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_activity\_merged`

ADD COLUMN DayOfWeek STRING;

ALTER TABLE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_activity\_merged`

ADD COLUMN SedentaryHours FLOAT64;

ALTER TABLE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_activity\_merged`

ADD COLUMN TotalActiveHours FLOAT64;

-- add required data to the above columns. For DayOfWeek, we have

UPDATE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_activity\_merged`

SET DayOfWeek = FORMAT\_DATE('%A', ActivityDate)

WHERE ActivityDate IS NOT NULL;

--sedentary hours column

UPDATE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_activity\_merged`

SET SedentaryHours = ROUND((SedentaryMinutes/60), 1)

WHERE SedentaryMinutes IS NOT NULL;

--total active hours update

UPDATE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_activity\_merged`

SET TotalActiveHours = ROUND((VeryActiveMinutes + FairlyActiveMinutes +

LightlyActiveMinutes + SedentaryMinutes)/60, 1)

WHERE TotalMinutes IS NOT NULL;

--also add columns for hours spent asleep, minutes taken to sleep in daily\_sleep\_merged. Update them with values

ALTER TABLE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_sleep\_merged`

ADD COLUMN HoursAsleep FLOAT64;

ALTER TABLE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_sleep\_merged`

ADD COLUMN MinutesTakenToSleep FLOAT64;

UPDATE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_sleep\_merged`

SET HoursAsleep = ROUND((TotalMinutesAsleep)/60, 1)

WHERE TotalMinutesAsleep IS NOT NULL;

UPDATE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_sleep\_merged`

SET MinutesTakenToSleep = TotalTimeInBed - TotalMinutesAsleep

WHERE TotalTimeInBed IS NOT NULL AND TotalMinutesAsleep IS NOT NULL;

--add a column to weight\_merged csv file to indicate whether the user is over or underweight, or whether they are in the healthy range.

ALTER TABLE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.weight\_merged`

ADD COLUMN BMIClass STRING;

UPDATE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.weight\_merged`

SET BMIClass = CASE

WHEN BMI > 25 THEN "Overweight"

WHEN BMI < 18.5 THEN "Underweight"

ELSE "Healthy" END

WHERE BMI IS NOT NULL;

--we can also remove the Fat column in weight, as it has no purpose in the analysis, since it contains null values

ALTER TABLE `da-capstone-bellabeat-399911.Bellabeat\_case\_files.weight\_merged`

DROP COLUMN IF EXISTS FAT;

-- In the daily\_sleep table, we see that the nap hours of some users is around 1 hour. So let us find the minimum and maximum to verify this observation and count how many users fall in each section

SELECT

MIN(HoursAsleep) AS MinHoursAsleep, MAX(HoursAsleep) AS MaxHoursAsleep,

FROM `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_sleep\_merged`

WHERE Id IS NOT NULL;

-- so the minimum number of hours is 1, and the maximum is 13.3. From the ReadMe using RStudio, I had already noted that the healthy sleeping hours for individuals is between 7-9 hours. Lets see how many users accomplish this

SELECT

COUNT(DISTINCT(Id)) AS NoOfUsersHealthySleeping,

COUNT(DISTINCT(Id)) AS NoOfUsersNapping,

COUNT(DISTINCT(Id)) As NoOfUsersOversleeping

FROM `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_sleep\_merged`

WHERE CASE

WHEN HoursAsleep < 5 THEN SUM(count\_min)

WHEN HoursAsleep > 9 THEN SUM(count\_max) END;

--from this, we understand that 19 users out of a total distinct count of 24 (79.2%) achieve healthy sleeping goals.

--we will check for the number of records that have either recorded a user nap time and who have slept for more than 9 hours

BEGIN

DECLARE count\_min, count\_max INT64;

SET count\_min = (

SELECT

COUNT(DISTINCT(Id)) As UsersNapping

FROM `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_sleep\_merged`

WHERE HoursAsleep < 4

);

SET count\_max = (

SELECT

COUNT(DISTINCT(Id)) AS UsersOversleeping

FROM `da-capstone-bellabeat-399911.Bellabeat\_case\_files.daily\_sleep\_merged`

WHERE HoursAsleep > 9

);

SELECT FORMAT('Number of users napping: %d and number of users oversleeping: %d', count\_min, count\_max) AS result;

END