-- View and examine tables

-- 940 records from Daily Activity table

SELECT \*

FROM `gda-course-4-332812.Capstone.DailyActivity` AS DailyActivity

-- Only 413 records from Sleep table

SELECT \*

FROM `gda-course-4-332812.Capstone.SleepDay` AS SleepDay

-- 1,325,580 records from Sleep table

SELECT \*

FROM `gda-course-4-332812.Capstone.MinuteMETs` AS MinuteMETS

-- Hourly intensities, calories, and steps each have 22,099 rows

SELECT \*

FROM `gda-course-4-332812.Capstone.HourlyIntensities` AS HourlyIntensities

SELECT \*

FROM `gda-course-4-332812.Capstone.HourlyCalories` AS HourlyCalories

SELECT \*

FROM `gda-course-4-332812.Capstone.HourlySteps` AS HourlySteps

-- Check for the number of unique participants

-- There are 33 unique participants in Daily Activities table

SELECT COUNT(DISTINCT Id) AS UserCount

FROM `gda-course-4-332812.Capstone.DailyActivity` AS DailyActivity

-- There are only 24 unique participants with sleep data

SELECT COUNT(DISTINCT Id) AS UserCount

FROM `gda-course-4-332812.Capstone.SleepDay` AS SleepDay

-- 33 unique participants from METs table.

SELECT COUNT(DISTINCT Id) AS UserCount

FROM `gda-course-4-332812.Capstone.MinuteMETs` AS MinuteMETS

-- Hourly tables each have 33 participants, matching with Daily Activity table.

SELECT COUNT(DISTINCT Id) AS UserCount

FROM `gda-course-4-332812.Capstone.HourlyIntensities` AS HourlyIntensities

SELECT COUNT(DISTINCT Id) AS UserCount

FROM `gda-course-4-332812.Capstone.HourlyCalories` AS HourlyCalories

SELECT COUNT(DISTINCT Id) AS UserCount

FROM `gda-course-4-332812.Capstone.HourlySteps` AS HourlySteps

-- HeartRate table only has 7 participants while WeightLogInfo has 8 participants, out of 33. These tables will not be used for this analysis due to this limitation.

-- Check for duplicates.

-- No duplicates found on DailyActivities table

SELECT Id,

ActivityDate,

TotalSteps,

TotalDistance,

TrackerDistance,

LoggedActivitiesDistance,

VeryActiveDistance,

ModeratelyActiveDistance,

LightActiveDistance,

SedentaryActiveDistance,

VeryActiveMinutes,

FairlyActiveMinutes,

LightlyActiveMinutes,

SedentaryMinutes,

Calories,

COUNT(\*) AS Dupes

FROM `gda-course-4-332812.Capstone.DailyActivity` AS DailyActivity

GROUP BY Id,

ActivityDate,

TotalSteps,

TotalDistance,

TrackerDistance,

LoggedActivitiesDistance,

VeryActiveDistance,

ModeratelyActiveDistance,

LightActiveDistance,

SedentaryActiveDistance,

VeryActiveMinutes,

FairlyActiveMinutes,

LightlyActiveMinutes,

SedentaryMinutes,

Calories

HAVING Dupes > 1

-- Found 3 duplicates from SleepDay table, removed those duplicates.

SELECT

Id,

SleepDay,

TotalSleepRecords,

TotalMinutesAsleep,

TotalTimeInBed,

COUNT(\*) AS Dupes

FROM `gda-course-4-332812.Capstone.SleepDay` AS SleepDay1

GROUP BY Id,

SleepDay,

TotalSleepRecords,

TotalMinutesAsleep,

TotalTimeInBed

HAVING Dupes = 1

ORDER BY Id,

SleepDay,

TotalSleepRecords,

TotalMinutesAsleep,

TotalTimeInBed

-- Created user no table to clearly identify unique participant

-- Merged Daily Activity, Sleep Day, and User No tables.

-- Nulls have been set to 0.

SELECT

DailyActivity.Id,

UserNumberTable.UserNo,

ActivityDate,

SUM(TotalSteps) AS Total\_Steps,

SUM(TotalDistance) AS Total\_Distance,

SUM(TrackerDistance) AS Total\_Tracker\_Distance,

SUM(LoggedActivitiesDistance) AS Total\_LoggedActivitiesDistance,

SUM(VeryActiveDistance) AS Total\_VeryActiveDistance,

SUM(ModeratelyActiveDistance) AS Total\_ModeratelyActiveDistance,

SUM(LightActiveDistance) AS Total\_LightActiveDistance,

SUM(SedentaryActiveDistance) AS Total\_SedentaryActiveDistance,

SUM(VeryActiveMinutes) AS Total\_VeryActiveMinutes,

SUM(FairlyActiveMinutes) AS Total\_FairlyActiveMinutes,

SUM(LightlyActiveMinutes) AS Total\_LightlyActiveMinutes,

SUM(SedentaryMinutes) AS Total\_SedentaryMinutes,

SUM(Calories) AS Total\_Calories,

IFNULL(SUM(TotalSleepRecords),0) AS Total\_SleepRecords,

IFNULL(SUM(TotalMinutesAsleep),0) AS Total\_MinutesAsleep,

IFNULL(SUM(TotalTimeInBed),0) AS Total\_TimeInBed,

FROM `gda-course-4-332812.Capstone.DailyActivity` AS DailyActivity

LEFT JOIN

(SELECT

Id,

SleepDay,

TotalSleepRecords,

TotalMinutesAsleep,

TotalTimeInBed,

COUNT(\*) AS Dupes

FROM `gda-course-4-332812.Capstone.SleepDay` AS SleepDay1

GROUP BY Id,

SleepDay,

TotalSleepRecords,

TotalMinutesAsleep,

TotalTimeInBed

HAVING Dupes = 1

ORDER BY Id,

SleepDay,

TotalSleepRecords,

TotalMinutesAsleep,

TotalTimeInBed) AS SleepDay1

ON DailyActivity.Id = SleepDay1.Id

AND DailyActivity.ActivityDate = SleepDay1.SleepDay

LEFT JOIN `gda-course-4-332812.Capstone.UserNumberTable` AS UserNumberTable

ON DailyActivity.Id = UserNumberTable.Id

GROUP BY DailyActivity.Id, UserNumberTable.UserNo, DailyActivity.ActivityDate

ORDER BY DailyActivity.Id, UserNumberTable.UserNo, DailyActivity.ActivityDate

--Hourly intensity data, extracted year, month, day, day name, hour

-- No duplicates found

SELECT HourlyIntensities.Id, HourlyIntensities.ActivityHour, SUM(TotalIntensity) AS Total\_Intensity, AVG(AverageIntensity) AS Avg\_Intensity,

EXTRACT(YEAR FROM HourlyIntensities.ActivityHour) AS Year,

EXTRACT(MONTH FROM HourlyIntensities.ActivityHour) AS Month,

EXTRACT(DAY FROM HourlyIntensities.ActivityHour) AS Day,

EXTRACT(DAYOFWEEK FROM HourlyIntensities.ActivityHour) AS DayName,

EXTRACT(Hour FROM HourlyIntensities.ActivityHour) AS Hour,

COUNT(\*) AS Duplicates

FROM `gda-course-4-332812.Capstone.HourlyIntensities` AS HourlyIntensities

GROUP BY HourlyIntensities.Id, HourlyIntensities.ActivityHour

ORDER BY HourlyIntensities.Id, HourlyIntensities.ActivityHour

--Hourly calories data, extracted year, month, day, day name, hour

-- No duplicates found

SELECT Id, ActivityHour, SUM(Calories) AS Total\_Calories,

EXTRACT(YEAR FROM ActivityHour) AS Year,

EXTRACT(MONTH FROM ActivityHour) AS Month,

EXTRACT(DAY FROM ActivityHour) AS Day,

EXTRACT(DAYOFWEEK FROM ActivityHour) AS DayName,

EXTRACT(Hour FROM ActivityHour) AS Hour,

COUNT(\*) AS Duplicates

FROM `gda-course-4-332812.Capstone.HourlyCalories` AS Hourly\_Calories

GROUP BY Id, ActivityHour

ORDER BY Id, ActivityHour

--Hourly steps data, extracted year, month, day, day name, hour

-- No duplicates found

SELECT Id, ActivityHour, SUM(StepTotal) AS Total\_Steps,

EXTRACT(YEAR FROM ActivityHour) AS Year,

EXTRACT(MONTH FROM ActivityHour) AS Month,

EXTRACT(DAY FROM ActivityHour) AS Day,

EXTRACT(DAYOFWEEK FROM ActivityHour) AS DayName,

EXTRACT(Hour FROM ActivityHour) AS Hour,

COUNT(\*) AS Duplicates

FROM `gda-course-4-332812.Capstone.HourlySteps` AS HourlySteps

GROUP BY Id, ActivityHour

ORDER BY Id, ActivityHour

--Merged 3 hourly tables into one

SELECT HourlyIntensities.Id, HourlyIntensities.ActivityHour, SUM(TotalIntensity) AS Total\_Intensity, AVG(AverageIntensity) AS Avg\_Intensity,

SUM(HourlyCalories.Total\_Calories) AS TotalCalories,

SUM(HourlySteps.Total\_Steps) AS TotalSteps,

EXTRACT(DATE FROM HourlyIntensities.ActivityHour) AS Date,

EXTRACT(YEAR FROM HourlyIntensities.ActivityHour) AS Year,

EXTRACT(MONTH FROM HourlyIntensities.ActivityHour) AS Month,

EXTRACT(DAY FROM HourlyIntensities.ActivityHour) AS Day,

FORMAT\_DATE('%A', DATE(HourlyIntensities.ActivityHour)) AS DayName,

CAST(EXTRACT(TIME FROM HourlyIntensities.ActivityHour) AS TIME) AS Time,

EXTRACT(HOUR FROM HourlyIntensities.ActivityHour) AS Hour

FROM `gda-course-4-332812.Capstone.HourlyIntensities` AS HourlyIntensities

LEFT JOIN

(SELECT Id, ActivityHour, SUM(Calories) AS Total\_Calories,

EXTRACT(YEAR FROM ActivityHour) AS Year,

EXTRACT(MONTH FROM ActivityHour) AS Month,

EXTRACT(DAY FROM ActivityHour) AS Day,

EXTRACT(DAYOFWEEK FROM ActivityHour) AS DayName,

EXTRACT(Hour FROM ActivityHour) AS Hour,

FROM `gda-course-4-332812.Capstone.HourlyCalories` AS Hourly\_Calories

GROUP BY Id, ActivityHour) AS HourlyCalories

ON HourlyIntensities.Id = HourlyCalories.Id

AND HourlyIntensities.ActivityHour = HourlyCalories.ActivityHour

LEFT JOIN

(SELECT Id, ActivityHour, SUM(StepTotal) AS Total\_Steps,

EXTRACT(YEAR FROM ActivityHour) AS Year,

EXTRACT(MONTH FROM ActivityHour) AS Month,

EXTRACT(DAY FROM ActivityHour) AS Day,

EXTRACT(DAY FROM ActivityHour) AS DayName,

EXTRACT(Hour FROM ActivityHour) AS Hour,

COUNT(\*) AS Duplicates

FROM `gda-course-4-332812.Capstone.HourlySteps` AS Hourly\_Steps

GROUP BY Id, ActivityHour

ORDER BY Id, ActivityHour) AS HourlySteps

ON HourlyIntensities.Id = HourlySteps.Id

AND HourlyIntensities.ActivityHour = HourlySteps.ActivityHour

GROUP BY HourlyIntensities.Id, HourlyIntensities.ActivityHour

ORDER BY HourlyIntensities.Id, HourlyIntensities.ActivityHour

--METs per minute, extracted year, month, day, day name, hour

-- No duplicates found

-- Aggregated into METs per day

WITH Minute\_METs AS (SELECT Id, SUM(METs) AS METs,

EXTRACT(DATE FROM ActivityMinute) AS Date,

EXTRACT(YEAR FROM ActivityMinute) AS Year,

EXTRACT(MONTH FROM ActivityMinute) AS Month,

EXTRACT(DAY FROM ActivityMinute) AS Day,

FORMAT\_DATE('%A', DATE(ActivityMinute)) AS DayName,

CAST(EXTRACT(TIME FROM ActivityMinute) AS TIME) AS Time,

EXTRACT(HOUR FROM ActivityMinute) AS Hour,

COUNT(\*) AS Duplicates

FROM `gda-course-4-332812.Capstone.MinuteMETs` AS MinuteMETS

GROUP BY Id, ActivityMinute

ORDER BY Id, ActivityMinute)

SELECT Id, Date, SUM(METs) AS Total\_METs, ROUND(AVG(METs), 2) AS Avg\_METs,

EXTRACT(YEAR FROM Date) AS Year,

EXTRACT(MONTH FROM Date) AS Month,

EXTRACT(DAY FROM Date) AS Day,

FORMAT\_DATE('%A', DATE(Date)) AS DayName

FROM Minute\_METs

GROUP BY Id, Date