**Points to note:**

-- 1. Work out average, minimum and maximum daily steps from dailyactivity

These calculations were executed, and I discovered that the minimum total steps = 0. The only possible reasoning being that total steps were not tracked. I discovered that 77/940 observations had 0 total steps and 15/33 (almost half) of the participants did not track this information at least once during the period of tracking. The average steps equalled to 7638 which is well below the recommendation of 10,000 a day according to The American Heart Association.

-- 2. What days are users the most active and what days are users the least active

Saturday is the day users are most active on averaging 8153 steps and Sunday is the day users were least active, averaging 6933 steps.

-- 3. How many calories do users burn on average

2304 calories are burned on average by users per day. According to performancelab, without exercise, people on average, burn 1800 calories. 2304 calories is a significant improvement by users. However, if users have a goal of losing weight, it is estimated that 3500 calories must be burned daily to burn a pound within a week (as suggested by menshealth.com).

-- 4. Determine if trackerdistance and totaldistance match and check the sum of varying distances to see if they match either trackerdistance/totaldistance

As there is not much context provided for these columns, I will not provide recommendations regarding distance as it shows that 14 records do not meet this criteria.

-- 5. Validation check for Total distance SUM for dailyactivity

251/940 rows show that the SUM of column does not equal TotalDistance  
256/940 rows show that the SUM of column does not equal TrackerDistance

As there are multiple observations which do not match the criteria, I decided not to use this information on distance to provide any insights as they may be inaccurate.

-- 6. Check average VeryActiveMinutes

21 minutes is the average VeryActiveMinutes of activity of users within this dataset per day. The CDC suggests a weekly 150 minutes of vigorous activity which equals on average 21 minutes a day. Users on average are meeting this requirement towards better physical health.

-- 7. Check average SedentaryMinutes

-- 8. How many users work out for 60 minutes or more if they have SedentaryMinutes of more than 991 minutes

Users on average spend 991 minutes sedentary per day. Let’s say the recommended 8 hours of this time is dedicated to sleep, that leaves around another 8.5 hours of being sedentary. A new study reported by CBC states that those who sit for 8 hours, whether that be for work or just sitting can cause health risks and that 60 minutes of activity a day could offset this. To avoid assumptions, we have a table with information on users and their sleep, I created a new table (dailyactivitywithsleep) and used INNER JOIN to join relevant columns in dailyactivity with relevant columns from sleepday.

-- 9. Create a new table detailing relevant columns from dailyactivity and sleepday

To avoid assumptions, we have a table with information on users and their sleep, I created a new table (dailyactivitywithsleep) and used INNER JOIN to join relevant columns in dailyactivity with relevant columns from sleepday resulting in 410 observations. Of these 410 observations, there are zero observations of users with >=991 SedentaryMinutes AND TotalMinutesAsleep >=480 AND VeryActiveMinutes >=60. This shows that users are not meeting the sleep and exercise recommendations as set by CBC for people who spend more than 8 hours sitting per day. They are not sleeping enough and not carrying out enough exercise.

-- 10. Create a new table hourlyintensities by joining hourlyintensities and hourlysteps tables together

A new table hourlyintensitiesandsteps was created to allow for easier and clearer analysis of both data.

-- 11. What hour is most popular for activity

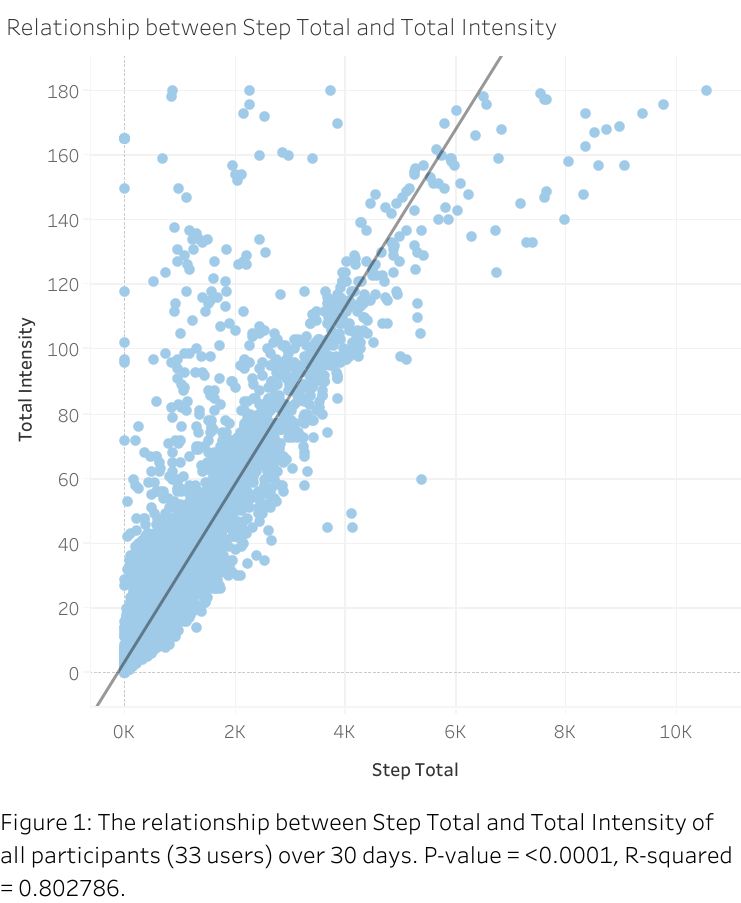
The average totalsteps was calculated to be 320 steps. To figure out which hours of the day were popular for activity. I used the average of 320 steps to filter out information for any observations recording less than 320 steps. 18:00 and 19:00 were the hours of the day where on average, users carried out the most activity. I also noted that 03:00 and 04:00 were the hours of least activity.

-- 12. Calculate average BMI of participants in weightloginfo

There is only information for 8 participants in weightloginfo as compared to 33 in dailyactivity so not all users have made use of this function. Information such as gender, age and height may have aided in a deeper analysis of this dataset. The average BMI is 25.2 which is overweight. BMI is also not an accurate measure for diagnosing body fat content. There may be different ethnicities as part of the study which hasn’t been clarified. Body fat distribution has been shown to vary between different ethnic/race groups as according to [health.clevelandclinic.org](https://health.clevelandclinic.org/is-bmi-accurate/). Only two observations recorded fat. The units were not confirmed and as it is such a small sample of the population of users, it is unnecessary to perform analysis on as it could inaccurately represent the data.

Using Tableau for initial exploration of data

I used Tableau Desktop 2022.4 to explore relationships between different variables that stood out to me whilst working with these datasets in MySQL workbench 8.0.



I noticed within the dataset ‘hourlyintensitiesandsteps’ that as the Total Intensity increased, the Step Total also increased for most observations. To confirm this for the entirety of the data, I created a scatterplot as shown in Figure 1. The relationship confirmed a positive correlation between both variables (also confirmed by low P-value = <0.0001 and high R-Squared = 0.802786. This shows that the more steps users are taking per day, the higher the total intensity.



Figure 2 shows that generally as a user’s total steps increased, the greater the number of calories they burned.

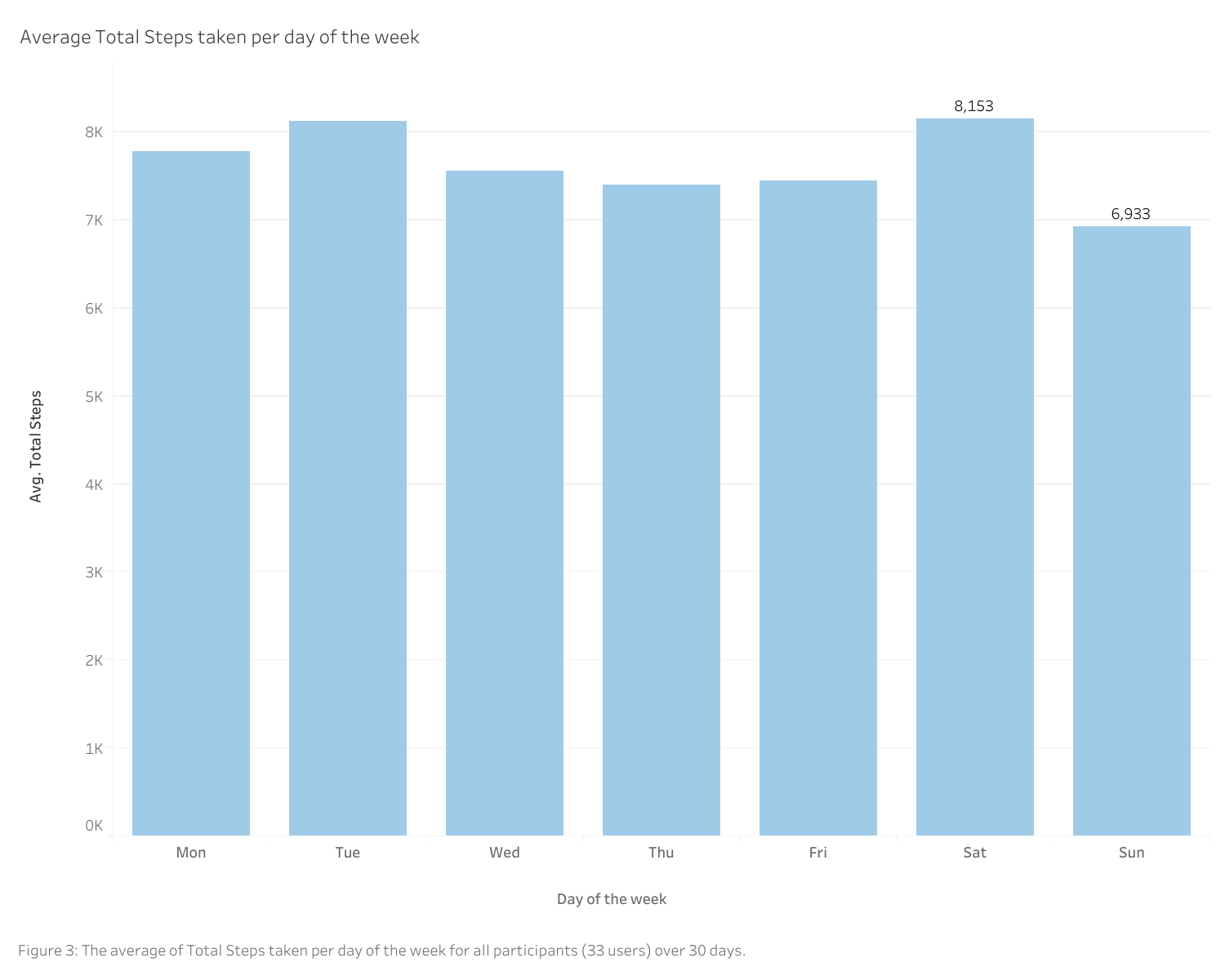
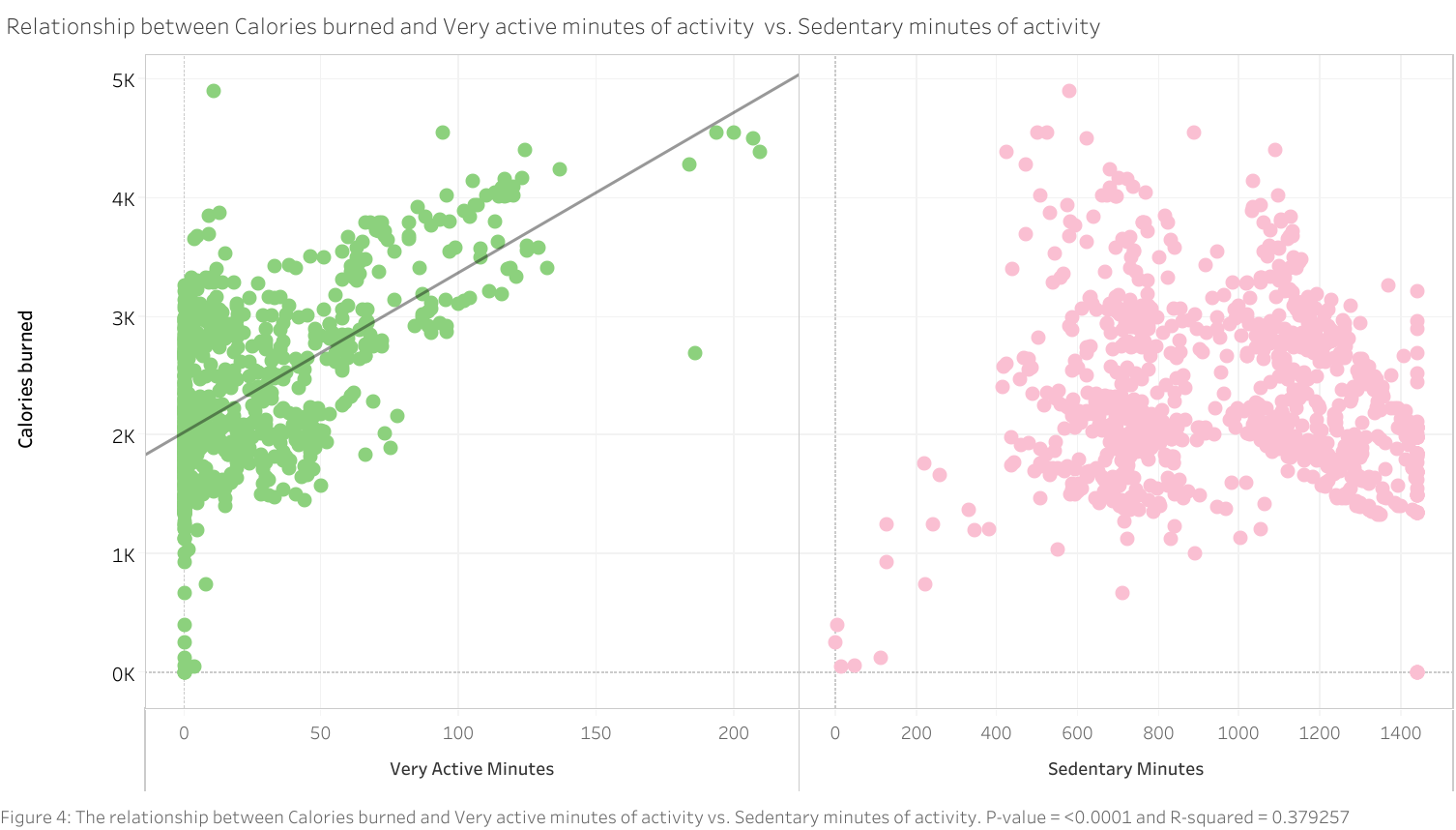


Figure 3 shows that Saturday followed by Tuesday are the days of most activity. With Saturday, users are averaging 8153 steps and Sunday being the day of least activity with users averaging 6933 steps.



In figure 4, there is a positive correlation between the number of calories burned and very active minutes of activity. The greater the activity in minutes, the greater the calories burned. Whereas there is not a correlation between minutes of being Sedentary and calories burned, the calories burned are maintained around the 1.5-2.5K mark for those sedentary for more than 600 minutes a day.

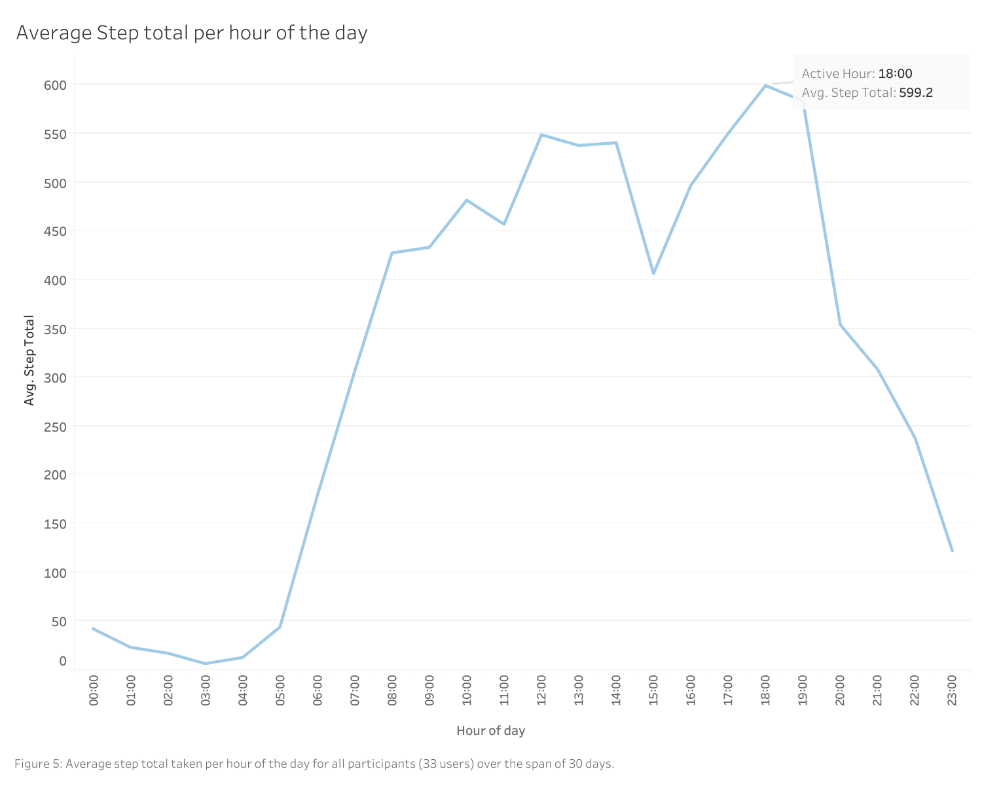


Figure 5 shows the trend of average total of steps taken over the 30 days at each hour. 6pm is the hour of greatest step total with 599 steps taken on average. Total steps fall from 8pm to 1am making this the period of least activity.

**SQL syntax and comments**

-- 1. Work out average daily steps from dailyactivity

SELECT avg(TotalSteps) FROM dailyactivity; -- the average steps being 7638 to 1.d.p.

SELECT max(TotalSteps) FROM dailyactivity; -- the max daily steps is 36019

SELECT min(TotalSteps) FROM dailyactivity; -- minimum shows as 0, so i will investigate this observation

-- There are 4 observations/days which were not tracked at all across 4 different individuals

-- There are also 77/940 records where 0 total steps were recorded.

SELECT \* FROM dailyactivity -- I inserted count to determine how many observations did not record their total steps

WHERE TotalSteps = 0

ORDER BY Calories ASC;

-- Determine how many participants didn't record totalsteps

SELECT count(DISTINCT(Id)) FROM dailyactivity -- 15/33 participants did not track one or more days of their dailyactivity.

WHERE TotalSteps = 0

SELECT count(DISTINCT(Id)) FROM dailyactivity -- 25/33 participants exceeded the recommended daily steps of 10,000 with 303 observations in total, cont.

WHERE TotalSteps > 10000;

SELECT count(\*) FROM dailyactivity -- cont. with 303/940 observations in total

WHERE TotalSteps > 10000;

-- 2. What days are users the most active and what days are users the least active

-- Create a new table with relevant dailyactivity information and add a column called 'day' to describe what day each observation is.

CREATE TABLE dailyactivitywithdayname

SELECT d.Id, d.ActivityDate, d.TotalSteps, d.TotalDistance, d.SedentaryMinutes, d.VeryActiveMinutes

FROM dailyactivity d;

ALTER TABLE dailyactivitywithdayname

ADD COLUMN day text;

UPDATE dailyactivitywithdayname

SET day=dayname(ActivityDate); -- column day has been updated with day name using Activity Date.

SELECT DAY, AVG(TotalSteps) FROM dailyactivitywithdayname -- On average Saturday is the most active day with users achieving 8152 steps

GROUP BY day -- On average Sunday is least active day with users achieving 6933 steps

ORDER BY AVG(Totalsteps) DESC;

-- 3. How many calories do users burn on average

SELECT avg(calories) FROM dailyactivity; -- 2304 calories are burned on average by users

-- 4. Determine if tracker distance and total distance match

-- As there is not much context provided for these columns, i will not provide recommendations regarding distance

SELECT \* FROM dailyactivity

WHERE trackerdistance <> totaldistance; -- this shows that are 14 records do not meet this criteria

-- 5. Validation check for Total distance SUM for dailyactivity

-- 251/940 rows show that the SUM of column does not equal TotalDistance

-- 256/940 rows show that the SUM of column does not equal TrackerDistance

SELECT COUNT(\*) FROM dailyactivity

WHERE (VeryActiveDistance + ModeratelyActiveDistance + LightActiveDistance) <> TotalDistance;

SELECT COUNT(\*) FROM dailyactivity

WHERE (VeryActiveDistance + ModeratelyActiveDistance + LightActiveDistance) <> TrackerDistance;

-- 6. Check average VeryActiveMinutes

SELECT avg(VeryActiveMinutes) FROM dailyactivity; -- 21 minutes is the average VeryActiveMinutes of activity

-- 7. Check average SedentaryMinutes

SELECT avg(SedentaryMinutes) FROM dailyactivity; -- 991 minutes is the average SedentaryMinutes of activity

-- 8. How many users work out for 60 minutes or more if they have SedentaryMinutes of more than 991 minutes

SELECT Count(\*) FROM dailyactivity

WHERE SedentaryMinutes >991 AND VeryActiveMinutes >60; -- 12/33 users on average carry out more than 60 minutes of activity if they are sedentary for more than 991 minutes per day.

-- 9. Create a new table dailyactivitywithsleep joining dailyactivity and sleepday together

CREATE TABLE dailyactivitywithsleep AS

SELECT da.Id,da.ActivityDate, da.TotalSteps,da.SedentaryMinutes,da.VeryActiveMinutes,da.Calories, sd.TotalMinutesAsleep,sd.TotalSleepRecords,sd.TotalTimeInBed

FROM dailyactivity da

INNER JOIN sleepday sd

ON da.Id = sd.Id AND da.ActivityDate = sd.SleepDate;

-- 10. Create a new table hourlyintensities by joining hourlyintensities and hourlysteps tables together

CREATE TABLE hourlyintensitiesandsteps

SELECT hi.Id, hi.ActiveDate, hi.ActiveHour, hi.TotalIntensity, hi.AverageIntensity, hs.StepTotal

FROM hourlyintensities hi

INNER JOIN hourlysteps hs

ON hi.Id = hs.Id AND hi.ActiveDate = hs.ActiveDate AND hi.ActiveHour = hs.ActiveHour;

-- 11. What hour is most popular for activity

SELECT avg(StepTotal) FROM hourlyintensitiesandsteps; -- average is 320 steps an hour for users

-- Calculate sum of step total for each hour to determine hour of most activity.

-- 18:00 is the hour of most activity by users with 03:00 being the hours of least activity

SELECT ActiveHour, sum(StepTotal)

FROM hourlyintensitiesandsteps

WHERE StepTotal >=320

GROUP BY ActiveHour

ORDER BY sum(StepTotal) DESC;

-- 12. Calculate average BMI of participants in weightloginfo

SELECT avg(BMI) FROM weightloginfo; -- 25.2 to 1.d.p is the average BMI of the 8 participants in weightloginfo