```
#  Data cleaning

 /*checking for how many users in the data - dailyActivity,
     30 users were specified in datasource but actual are 33 */

SELECT distinct(ID) FROM `bellabeatproject-363419.bellabeat.dailyActivity` LIMIT 100

#how many distinct users in dailyCalories

select distinct(ID) from `bellabeatproject-363419.bellabeat.dailyCalories` limit 100

#output 33 users

#how many distinct users in minuteMET

select distinct(ID) from `bellabeatproject-363419.bellabeat.minuteMET` limit 100

#output 33 users


#how many distinct users in sleepDay

select distinct(ID) from `bellabeatproject-363419.bellabeat.sleepDay` limit 100

#output 24 users

#how many distinct users in weightBMI

select distinct(ID) from `bellabeatproject-363419.bellabeat.weightBMI` limit 100

#output 8 users are have their weight record

#checking integrity of DATA in dailyActivity Table
# Check Total distance if it is correct

select TotalDistance,
round((SedentaryActiveDistance+LightActiveDistance+ModeratelyActiveDistance+VeryActive
Distance),2) as check_total
from `bellabeatproject-363419.bellabeat.dailyActivity`

#output -  data is correct , used round() function to show only 2 decimal place
```

```sql
--Data Exploration

/* Relation between Heart rate and steps */
select distinct h.Id,
  round(avg(h.value),2) as avg_heartRate,
  round(avg(d.TotalSteps),2) as avg_steps,
  round(avg(d.VeryActiveDistance),2) as avg_veryActiveDistance,
  round(avg(d.VeryActiveMinutes+d.FairlyActiveMinutes+d.LightlyActiveMinutes),2) as
avg_TotalActiveMinutes
  from bellabeat.heartRate h
  inner join bellabeat.dailyActivity d
  on h.Id = d.Id
  group by h.Id
  order by h.Id

#saved the result of this query in CSV format heartRate_vs_TotalActiveMins

/* BMI and weight vs totalsteps, totalDistance, ActiveMins, Calories */

select distinct w.Id,
  round(avg(w.WeightKg),2) as avg_Weight_kg,
  round(avg(w.BMI),2) as avg_BMI,
  round(avg(d.TotalSteps),2) as avg_totalSteps,
  round(avg(d.TotalDistance),2) as avg_totalDistance,
  round(avg(d.VeryActiveMinutes),2) as avg_veryActiveMins,
  round(avg(d.Calories),2) as avg_calories
  from bellabeat.weightBMI w
  inner join bellabeat.dailyActivity d
  on w.Id = d.Id
  group by w.Id
  order by w.Id
  #output - saved the result in CSV format BMI_vs_totalStepsCalories
```

```sql
-- write a query to extract weekday from the date
--then save results in new table for further calculations table name - weekday_data

select Id,ActivityDate, TotalDistance,TotalSteps,VeryActiveDistance,
VeryActiveMinutes,Calories,extract(dayofweek from ActivityDate) as weekday_number,
  case
    when extract(dayofweek from ActivityDate) = 1 then 'Sunday'
    when extract(dayofweek from ActivityDate) = 2 then 'Monday'
    when extract(dayofweek from ActivityDate) = 3 then 'Tuesday'
    when extract(dayofweek from ActivityDate) = 4 then 'Wednesday'
    when extract(dayofweek from ActivityDate) = 5 then 'Thursday'
    when extract(dayofweek from ActivityDate) = 6 then 'Friday'
    when extract(dayofweek from ActivityDate) = 7 then 'Saturday'
    else 'Invalid Input'
  end as weekday
from `bellabeatproject-363419.bellabeat.dailyActivity`



 --check if users are more active on weekends
 --save results in new summary table avg_weekends_activity

select round(avg(TotalSteps),2) as avg_TotalSteps, round(avg(Calories),2) as
avg_Calories,
  round(avg(TotalDistance),2) as avg_TotalDistance,
  round(avg(VeryActiveDistance),2) as avg_VeryActiveDistance,
  round(avg(VeryActiveMinutes),2) as avg_VeryActiveMinutes, weekday
  from `bellabeatproject-363419.bellabeat.weekday_data`
  where weekday_number= 1 or weekday_number =7
  group by weekday

--check if users are more active on weekdays
--save results in new summary table avg_weekdays_activity

select round(avg(TotalSteps),2) as avg_TotalSteps, round(avg(Calories),2) as
avg_Calories,
  round(avg(TotalDistance),2) as avg_TotalDistance,
  round(avg(VeryActiveDistance),2) as avg_VeryActiveDistance,
  round(avg(VeryActiveMinutes),2) as avg_VeryActiveMinutes,weekday
  from `bellabeatproject-363419.bellabeat.weekday_data`
  where weekday_number in (2,3,4,5,6)
```

```
  group by weekday



--check if users have more sedentary minutes on weekends
--save results in avg_sedentary_min_weekends

select round(avg(SedentaryMinutes)) as avg_sedentaryMinutes,
  extract(dayofweek from ActivityDate) as num_of_day
  from bellabeat.dailyActivity
  where (extract(dayofweek from ActivityDate)) in (1,7)
  group by num_of_day



--check if users have more sedentary minutes on weekdays,display maximun sedentary
minutes at top
--store results in avg_sedentaryMins_weekday

select round(avg(SedentaryMinutes)) as avg_sedentaryMinutes,(extract(dayofweek from
ActivityDate)) as num
  from bellabeat.dailyActivity
  where (extract(dayofweek from ActivityDate)) in (2,3,4,5,6)
  group by num
  order by avg(SedentaryMinutes) desc
```