Códigos utilizados no estudo

RStudio

```
# Importa o arquivo csv para o RStudio

> library(readr)

> BellabeatData <- read_csv("Matheus/Estudo de caso/BellabeatData.CSV",

+     col_types = cols(Id = col_character(),

+       `Activity Date` = col_date(format = "%d/%m/%Y"),

+       `Total Steps` = col_integer(), `Fairly Active Minutes` = col_integer(),

+       `Lightly Active Minutes` = col_integer(),

+       `Sedentary Minutes` = col_integer(),

+       `Calories` = col_integer()))
```

```
# Importa o arquivo csv para o RStudio

 Sleep <- read.csv("Matheus/Estudo de caso/sleepDay_organized.csv",

  col_types = cols(Id = col_character(),

               SleepDay = col_date(format = "%m/%d/%Y"),

               TotalSleepRecords = col_integer(),

               TotalMinutesAsleep = col_integer(),

               TotalTimeInBed = col_integer()))
```

```
# Importa o arquivo csv para o RStudio

Weight <- read_csv("Matheus/Estudo de caso/weight.csv",

  col_types = cols(

    Id = col_character(),

    Date = col_date(format= "%d/%m/%Y"),
    WeightKg = col_double(),
    WeightPounds = col_double(),
    Fat = col_integer(),
    BMI = col_double(),
    IsManualReport = col_guess(),
  ))


# Cria um backup da tabela de dados
BellabeatDataBCK <- BellabeatData

# Exclui linhas com valores nulos
BellabeatData[na.omit(BellabeatData)] <-

# Verifica letras entre em colunas numéricas
linha_com_letra <- grep("[A-Za-z]", BellabeatData$`Activity Date`)

#Remove duplicatas
BellabeatData <- distinct(BellabeatData)

# Remove espaços (trim)
BellabeatData <- BellabeatData %>%
mutate_all(trimws)

# Extrai a tabela de dados do RStudio no formato CSV
write.csv(Sleep, "Matheus/Estudo de caso/Sleep.csv", row.names = FALSE)

# Extrai a tabela de dados do RStudio no formato CSV
write.csv(BellabeatData, "Matheus/Estudo de caso/Merged.csv", row.names = FALSE)
```

SQL (Big Query)

```
-- Retorna a porcentagem de pessoas consideradas fisicamente ativas
SELECT
  COUNT(DISTINCT id)/33*100
FROM(
  SELECT
    Id,
    SUM(Fairly_Active_Minutes) AS Total_Fairly,
    SUM(Very_Active_Minutes) AS Total_Very,
  FROM
    `spartan-vertex-394720.Bellabeat.Merged`
  GROUP BY
    Id
)
WHERE
  Total_Fairly > 600
  OR Total_Very > 300


--Retorna a porcentagem de pessoas que utilizaram o monitoramento de corrida
SELECT
  COUNT(id)/33*100 AS Porcentage
FROM(
  SELECT
    id,
    COUNT(Total_Distance) AS Times_used
  FROM
  `spartan-vertex-394720.Bellabeat.Merged`
  WHERE
    Total_Distance > 5
  GROUP BY
    Id
)
WHERE
  Times_used > 15

--Retorna a porcentagem de pessoas que utilizaram o monitoramento de sono
SELECT
  COUNT(DISTINCT id)/33*100 AS Porcentage
FROM(
  SELECT
    id,
    COUNT(SleepDay) AS Times_Used
  FROM
    `spartan-vertex-394720.Bellabeat.Sleep`
  GROUP BY
    Id
)
WHERE
  Times_Used > 15


--Retorna a porcentagem de pessoas que utilizaram o registro de peso
SELECT
  COUNT(DISTINCT Id)/33*100 AS Porcentage
FROM
  `spartan-vertex-394720.Bellabeat.Weight`



--Retorna quantas pessoas de m2 também estão em m1
SELECT COUNT(DISTINCT id)
FROM (
  SELECT DISTINCT m1.id
  FROM (
    SELECT
      Id,
      SUM(Fairly_Active_Minutes) AS Total_Fairly,
      SUM(Very_Active_Minutes) AS Total_Very
    FROM `spartan-vertex-394720.Bellabeat.Merged`
    GROUP BY Id
  ) AS m1
  WHERE
    (m1.Total_Fairly > 600 OR m1.Total_Very > 300)
    AND m1.id IN (
      SELECT DISTINCT m2.id
      FROM (
        SELECT
          id,
          COUNT(Total_Distance) AS Times_used
        FROM `spartan-vertex-394720.Bellabeat.Merged`
        WHERE Total_Distance > 5
        GROUP BY Id
      ) AS m2
      WHERE m2.Times_used > 15
    )
)
```