**SQL QUERIES FITBIT CASE STUDY**

--COUNT OF PARTECIPANTS BY DISTINCT IDs FROM EACH DATASET(6)

--1

**SELECT** **COUNT**(**DISTINCT** Id) **AS** Number\_Partecipants

**FROM** "1\_Daily\_Activity\_CSV" dac; --33

--2

**SELECT** **COUNT**(**DISTINCT** Id) **AS** Number\_Partecipants

**FROM** "2\_Daily\_Sleep\_CSV" dsc; --24

--3

**SELECT** **COUNT**(**DISTINCT** Id) **AS** Number\_Partecipants

**FROM** "3\_Hourly\_Calories\_CSV" hcc; --33

--4

**SELECT** **COUNT**(**DISTINCT** Id) **AS** Number\_Partecipants

**FROM** "4\_Hourly\_Intensity\_CSV" hic; --33

--5

**SELECT** **COUNT**(**DISTINCT** Id) **AS** Number\_Partecipants

**FROM** "5\_Hourly\_Steps\_CSV" hsc; --33

--6

**SELECT** **COUNT**(**DISTINCT** Id) **AS** Number\_Partecipants

**FROM** "6\_Weight\_Log\_CSV" wlc; --8

--THE DATASET 6\_Weight\_Log\_CSV DOES NOT INCLUDE ENOUGH DATA, HENCE IT WON'T BE USED.

--TIMES EACH PARTECIPANTS USED THE FITBIT

**SELECT** Id, **COUNT**(**DISTINCT** ActivityDate) **AS** DateCount

**FROM** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id;

--COUNT OF DateCount TO SEE THE FREQUENCY

**SELECT** DateCount, **COUNT**(DateCount) **AS** CountOfCounts

**FROM**

(**SELECT** Id, **COUNT**(**DISTINCT** ActivityDate) **AS** DateCount

**FROM** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id

) **AS** subquery

**GROUP** **BY** DateCount

**ORDER** **BY** CountOfCounts **DESC**;

--DIVIDING USERS IN 3 CATEGORIES BASED ON THE USAGE

**SELECT** Id,

**COUNT**(Id) **AS** Logged\_Times,

**CASE**

**WHEN** **COUNT**(Id) **BETWEEN** 25 **AND** 31 **THEN** 'Active Usage'

**WHEN** **COUNT**(Id) **BETWEEN** 15 **and** 24 **THEN** 'Moderate Usage'

**WHEN** **COUNT**(Id) **BETWEEN** 0 **and** 14 **THEN** 'Light Usage'

**END** Usage\_Type

**FROM** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id;

--HOW MANY OF EACH USAGE TYPE

**SELECT** Logged\_Times, Usage\_Type, **COUNT**(Usage\_Type) **AS** Count\_Usage\_Type

**FROM**

(**SELECT** Id,

**COUNT**(Id) **AS** Logged\_Times,

**CASE**

**WHEN** **COUNT**(Id) **BETWEEN** 25 **AND** 31 **THEN** "Active Usage"

**WHEN** **COUNT**(Id) **BETWEEN** 15 **and** 24 **THEN** "Moderate Usage"

**WHEN** **COUNT**(Id) **BETWEEN** 0 **and** 14 **THEN** "Light Usage"

**END** Usage\_Type

**FROM** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id)

**AS** subquery

**GROUP** **BY** Usage\_Type

**ORDER** **BY** **COUNT**(Usage\_Type) **DESC**;

--EACH USER WITH TOTAL STEPS, DISTANCE AND CALORIES COUNT

**SELECT** Id, **SUM**(TotalSteps) **AS** Sum\_Steps, **SUM**(TotalDistance) **AS** Sum\_DIstance, **SUM**(Calories) **AS** Sum\_Calories

**FROM** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id

**ORDER** **BY** **SUM**(TotalSteps) **DESC**;

--AVG OF TOTAL STEPS, TOTAL DISTANCE AND CALORIES FOR EACH USER

**SELECT** Id, **AVG**(TotalSteps) **AS** Avg\_Steps, **AVG**(TotalDistance) **AS** Avg\_DIstance, **AVG**(Calories) **AS** Avg\_Calories

**FROM** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id

**ORDER** **BY** **AVG**(TotalSteps) **DESC**;

--AVG TIME FOR ACTIVITY LEVEL(VERY ACTIVE,FAIRLY ACTIVE LIGHTLY ACTIVE, SEDENTARY) FOR EACH USER PER DAY

**SELECT** Id,

**AVG**(VeryActiveMinutes) **AS** Avg\_Very\_Active\_Minutes,

**AVG**(FairlyActiveMinutes) **AS** Avg\_Fairly\_Active\_Minutes,

**AVG**(LightlyActiveMinutes) **AS** Avg\_Lightly\_Active\_Minutes,

**AVG**(SedentaryMinutes) **AS** Avg\_Sedentary\_Minutes

**From** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id;

--DO THE USERS HIT THE CDC RECOMMANDED TIME OF PHYSICAL ACTIVITY EACH WEEK(150 MINS)

**SELECT** Id, (**AVG**(VeryActiveMinutes) + **AVG**(FairlyActiveMinutes))\*7 **AS** Avg\_Active\_Minutes\_Week,

**CASE**

**WHEN** (**AVG**(VeryActiveMinutes) + **AVG**(FairlyActiveMinutes))\*7 >= 150

**THEN** "YES"

**ELSE** "NO"

**END** CDC\_Recommendation\_Met

**FROM** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id

**ORDER** **BY** Avg\_Active\_Minutes\_Week **DESC**;

--DO THE USERS HIT THE RECOMMANDED STEPS PER DAY

**SELECT** Id, **AVG**(TotalSteps) **AS** Avg\_Steps,

**CASE**

**WHEN** **AVG**(TotalSteps) < 3000

**THEN** "Inactive"

**WHEN** **AVG**(TotalSteps) **BETWEEN** 3000 **AND** 4999

**THEN** "Low Active User"

**WHEN** **AVG**(TotalSteps) **BETWEEN** 5000 **AND** 7999

**THEN** "Average Active User"

**WHEN** **AVG**(TotalSteps) **BETWEEN** 8000 **AND** 12000

**THEN** "Active User"

**ELSE** "Very Active User"

**END** "User\_Type"

**FROM** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id

**ORDER** **BY** **AVG**(TotalSteps) **DESC**;

--HOW MANY OF EACH USER TYPE

**SELECT** User\_Type, **COUNT**(User\_Type) **AS** Count\_User\_Type

**FROM**

(**SELECT** Id, **AVG**(TotalSteps) **AS** Avg\_Steps,

**CASE**

**WHEN** **AVG**(TotalSteps) < 3000

**THEN** "Inactive"

**WHEN** **AVG**(TotalSteps) **BETWEEN** 3000 **AND** 4999

**THEN** "Low Active User"

**WHEN** **AVG**(TotalSteps) **BETWEEN** 5000 **AND** 7999

**THEN** "Average Active User"

**WHEN** **AVG**(TotalSteps) **BETWEEN** 8000 **AND** 12000

**THEN** "Active User"

**ELSE** "Very Active User"

**END** "User\_Type"

**FROM** "1\_Daily\_Activity\_CSV" dac

**GROUP** **BY** Id)

**AS** subquery

**GROUP** **BY** User\_Type

**ORDER** **BY** **COUNT**(User\_Type) **DESC**;

--AVERAGE STEPS BY HOURS FOR USERS

**SELECT** ActivityHour, **AVG**(StepTotal)

**FROM** "5\_Hourly\_Steps\_CSV" hsc

**GROUP** **BY** ActivityHour

**ORDER** **BY** **AVG**(StepTotal) **DESC**;

--TOTAL STEPS BY HOURS FOR USERS

**SELECT** ActivityHour, **SUM**(StepTotal)

**FROM** "5\_Hourly\_Steps\_CSV" hsc

**GROUP** **BY** ActivityHour

**ORDER** **BY** **SUM**(StepTotal) **DESC**;

--SLEEP in RELATION TO STEPS AND CALORIES FOR EACH USER

**SELECT** dsc.Id, **AVG**(TotalMinutesAsleep) **AS** Avg\_Sleep\_Time, **AVG**(TotalSteps), **AVG**(Calories)

**FROM** "2\_Daily\_Sleep\_CSV" dsc

**INNER** **JOIN** "1\_Daily\_Activity\_CSV" dac

**ON** dsc.Id=dac.Id

**GROUP** **BY** dsc.Id

**ORDER** **BY** Avg\_Sleep\_Time **DESC**;

--MOST ACTIVE HOUR FOR CALORIES SPENT

**SELECT** ActivityHour, **AVG**(Calories) **AS** AverageCalories

**FROM** "3\_Hourly\_Calories\_CSV" hcc

**GROUP** **BY** ActivityHour

**ORDER** **BY** AverageCalories **DESC**;