Homework | module 2 > week 6 > day 15

Topics covered: *Google BigQuery, WHERE, CASE-WHEN*

Standard Exercise:

Go to the Google BigQuery Sandbox environment and search for the `bigquery-public-data.ml_datasets.census_adult_income` dataset among the public datasets. Open a query editor and try to answer the following questions *using only the clauses and statements seen in class so far*.

1. Check the range of the age variable, it goes from 17 to 90 years old
   Usare min(age)
   max(age)

; using a CASE-WHEN statement, create a new variable that groups the data in 5 age brackets: "<20", "20-39", "40-59", "60-79", ">=80"

```
SELECT
  CASE
  WHEN age < 20 THEN 'young'
  WHEN (age>= 20 AND  age <= 39) THEN '20-39'
  WHEN (age>= 40 AND age <= 59) THEN '40-59'
  WHEN (age>= 60 AND age <= 79) THEN '60-79'
  WHEN age >=80  THEN '80'
 ELSE 'not_found'
 END as age_person
FROM `bigquery-public-data.ml_datasets.census_adult_income`
 LIMIT 1000;
```

2. To make sure your statement is working correctly, combine it with a WHERE clause to see if the brackets contain the correct ages (for instance, you could include only age <20 and check that only the bracket "<20" is returned).

```
SELECT age, sex,
 CASE
 WHEN age < 20 THEN 'young'
 WHEN (age>= 20 AND  age <= 39) THEN '20-39'
 WHEN (age>= 40 AND age <= 59) THEN '40-59'
 WHEN (age>= 60 AND age <= 79) THEN '60-79'
 WHEN age >=80  THEN '80'
ELSE 'not_found'
```

```
   END as age_person
 FROM `bigquery-public-data.ml_datasets.census_adult_income`
 where age < 20
  LIMIT 1000;
```

3. How many white males are there with a Doctorate in the United States?

```
SELECT
   Count(*) as n_male_with_doctorate,
 FROM `bigquery-public-data.ml_datasets.census_adult_income`
WHERE upper(trim(sex)) = 'MALE' and UPPER(trim(race))= 'WHITE' and
UPPER(trim(education))= 'Masters' and trim(native_country) ='United_States'
```

4. Is it more or less than the number of white females with a Doctorate in the United States?

```
SELECT
   Count(*) as n_female_with_doctorate,
 FROM `bigquery-public-data.ml_datasets.census_adult_income`
WHERE upper(trim(sex)) = 'FEMALE' and UPPER(trim(race))= 'WHITE' and
UPPER(trim(education))= 'Masters' and trim(native_country) ='United_States'
```

5. In the previous question, you should have observed that, *in absolute terms*, there are almost four times as many males with a doctorate compared to females. But is it really so? Well, in absolute terms it is, however, to get a

more accurate picture of our population, we should compare those values to the proportion of total males and females in the population.
Find how many males and females there are in the total population.

```
SELECT
   Count(*) as n_female,
 FROM `bigquery-public-data.ml_datasets.census_adult_income`
WHERE upper(trim(sex)) = 'FEMALE'; = 10771

SELECT
   Count(*) as n_male,
 FROM `bigquery-public-data.ml_datasets.census_adult_income`
WHERE upper(trim(sex)) = 'MALE' ; = 21790
```